

The Geometry of Differential Privacy: The Sparse and Approximate Cases

Aleksandar Nikolov*

Kunal Talwar†

Li Zhang‡

Abstract

In this work, we study trade-offs between accuracy and privacy in the context of linear queries over histograms. This is a rich class of queries that includes contingency tables and range queries, and has been a focus of a long line of work [BLR08, RR10, DRV10, HT10, HR10, LHR⁺10, BDKT12]. For a given set of d linear queries over a database $x \in \mathbb{R}^N$, we seek to find the differentially private mechanism that has the minimum mean squared error. For pure differential privacy, [HT10, BDKT12] give an $O(\log^2 d)$ approximation to the optimal mechanism. Our first contribution is to give an $O(\log^2 d)$ approximation guarantee for the case of (ϵ, δ) -differential privacy. Our mechanism is simple, efficient and adds carefully chosen correlated Gaussian noise to the answers. We prove its approximation guarantee relative to the *hereditary discrepancy* lower bound of [MN12], using tools from convex geometry.

We next consider this question in the case when the number of queries exceeds the number of individuals in the database, i.e. when $d > n \triangleq \|x\|_1$. The lower bounds used in the previous approximation algorithm no longer apply, and in fact better mechanisms are known in this setting [BLR08, RR10, HR10, GHRU11, GRU12]. Our second main contribution is to give an (ϵ, δ) -differentially private mechanism that for a given query set A and an upper bound n on $\|x\|_1$, has mean squared error within $\text{polylog}(d, N)$ of the optimal for A and n . This approximation is achieved by coupling the Gaussian noise addition approach with linear regression over the ℓ_1 ball. Additionally, we show a similar polylogarithmic approximation guarantee for the best ϵ -differentially private mechanism in this sparse setting. Our work also shows that for arbitrary counting queries, i.e. A with entries in $\{0, 1\}$, there is an ϵ -differentially private mechanism with expected error $\tilde{O}(\sqrt{n})$ per query, improving on the $\tilde{O}(n^{\frac{2}{3}})$ bound of [BLR08], and matching the lower bound implied by [DN03] up to logarithmic factors.

The connection between hereditary discrepancy and the privacy mechanism enables us to derive the first polylogarithmic approximation to the hereditary discrepancy of a matrix A .

1 Introduction

Differential privacy [DMNS06] is a recent privacy definition that has quickly become the standard notion of privacy in statistical databases. Informally, a mechanism (a randomized function on databases) satisfies differential privacy if the distribution of the outcome of the mechanism does not change noticeably when one individual's input to the database is changed. Privacy is measured by how small this change must be: an ϵ -differentially private (ϵ -DP) mechanism \mathcal{M} satisfies $\Pr[\mathcal{M}(x) \in S] \leq \exp(\epsilon) \Pr[\mathcal{M}(x') \in S]$ for any pair x, x' of neighboring databases, and for any measurable subset S of the range. A relaxation of this definition is *approximate differential privacy*. A mechanism \mathcal{M} is (ϵ, δ) -differentially private ((ϵ, δ) -DP) if $\Pr[\mathcal{M}(x) \in S] \leq \exp(\epsilon) \Pr[\mathcal{M}(x') \in S] + \delta$ with x, x', S as before. Here δ is thought of as negligible in the size of the database. Both these definitions satisfy several desirable properties such as composability, and are resistant to post-processing of the output of the mechanism.

In recent years, a large body of research has shown that this strong privacy definition still allows for very accurate analyses of statistical databases. At the same time, answering a large number of adversarially chosen queries accurately is inherently impossible with any semblance of privacy. Indeed Dinur and Nissim [DN03] show that answering $\tilde{O}(d)$ random subset sums (\tilde{O} hides polylogarithmic factors in $N, d, 1/\delta$.) of a set of d bits with (per query) error $o(\sqrt{d})$ allows an attacker to reconstruct (an arbitrarily good approximation to) all the private information. Thus there is an inherent trade-off between privacy and accuracy when answering a large number of queries. In this work, we study this trade-off in the context of counting queries, and more generally linear queries.

We think of the database as being given by a multiset of database rows, one for each individual. We will let N denote the size of the universe that these rows come from, and we will denote by n the number of individuals in

*Department of Computer Science, Rutgers University, Piscataway, NJ 08854. This work was done while the author was at Microsoft Research SVC.

†Microsoft Research SVC, Mountain View, CA 94043.

‡Microsoft Research SVC, Mountain View, CA 94043.

the database. We can represent the database as its histogram $x \in \mathbb{R}^N$ with x_i denoting the number of occurrences of the i th element of the universe. Thus x would in fact be a vector of non-negative integers with $\|x\|_1 = n$. We will be concerned with reporting reasonably accurate answers to a given set of d linear queries over this histogram x . This set of queries can naturally be represented by a matrix $A \in \mathbb{R}^{d \times N}$ with the vector $Ax \in \mathbb{R}^d$ giving the correct answers to the queries. When $A \in \{0, 1\}^{d \times N}$, we call such queries counting queries. We are interested in the (practical) regime where $N \gg d \gg n$, although our results hold for all settings of the parameters.

A differentially private mechanism will return a noisy answer to the query A and, in this work, we measure the performance of the mechanisms in terms of its worst case total expected squared error. Suppose that $X \subseteq \mathbb{R}^N$ is the set of all possible databases. The error of a mechanism \mathcal{M} is defined as $\text{err}_{\mathcal{M}}(A, X) = \max_{x \in X} \mathbb{E}[\|\mathcal{M}(x) - Ax\|_2^2]$. Here the expectation is taken over the internal coin tosses of the mechanism itself, and we look at the worst case of this expected squared error over all the databases in X . Unless stated otherwise, the error of the mechanism will refer to this worst case expected ℓ_2^2 error. Phrased thus, the Gaussian noise mechanism of Dwork et al. [DKM⁺06a] gives error at most $O(d^2)$ for any counting query and guarantees (ε, δ) -DP¹ over all the databases, i.e. $X = \mathbb{R}^N$. Moreover, the aforementioned lower bounds imply that there exist counting queries for which this bound can not be improved. For ε -DP, Hardt and Talwar [HT10] gave a mechanism with error $O(d^2 \log \frac{N}{d})$ and showed that this is the best possible for random counting queries. Thus the worst case accuracy for counting queries is fairly well-understood in this measure.

Specific sets of counting queries of interest can however admit much better mechanisms than adversarially chosen queries for which the lower bounds are shown. Indeed several classes of specific queries have attracted attention. Some, such as range queries, are “easier”, and asymptotically better mechanisms can be designed for them. Others, such as constant dimensional contingency tables, are nearly as hard as general counting queries, and asymptotically better mechanisms can be ruled out in some ranges of the parameters. These query-specific upper bounds are usually proved by carefully exploiting the structure of the query, and query-specific lower bounds have been proved by reconstruction attacks that exploit a lower bound on the smallest singular value of an appropriately chosen A [DN03, DMT07, DY08, KRSU10, De12, KRS13]. It is natural to address this question in a competitive analysis framework: can we design an efficient algorithm that given any query A , computes (even approximately) the minimum error differentially private mechanism for A ?

Hardt and Talwar [HT10] answered this question in the affirmative for ε -DP mechanisms, and gave a mechanism that has error within factor $O(\log^3 d)$ of the optimal assuming a conjecture from convex geometry known as the hyperplane conjecture or the slicing conjecture. Bhaskara et al. [BDKT12] removed the dependence on the hyperplane conjecture and improved the approximation ratio to $O(\log^2 d)$. Can relaxing the privacy requirement to (ε, δ) -DP help with accuracy? In many settings, (ε, δ) -DP mechanisms can be simpler and more accurate than the best known ε -DP mechanisms. This motivates the first question we address.

Question 1 *Given A , can we efficiently approximate the optimal error (ε, δ) -DP mechanism for it?*

Hardt and Talwar [HT10] showed that for some A , the lower bound for ε -DP mechanism can be $\Omega(\log N)$ larger than known (ε, δ) -DP mechanisms. For non-linear Lipschitz queries, De [De12] showed that this gap can be as large as $\Omega(\sqrt{d})$ (even when $N = d$). This leads us to ask:

Question 2 *How large can the gap between the optimal ε -DP mechanism and the optimal (ε, δ) -DP mechanism be for linear queries?*

When the databases are sparse, e.g. when $\|x\|_1 \leq n \ll d$, one may obtain better mechanisms. Blum, Ligett and Roth [BLR08] gave an ε -DP mechanism that can answer any set of d counting queries with error $\tilde{O}(dn^{\frac{4}{3}})$. A series of subsequent works [DNR⁺09, DRV10, RR10, HR10, GHRU11, HLM12, GRU12] led to (ε, δ) -DP mechanisms that have error only $\tilde{O}(dn)$. Thus when $n < d$, the lower bound of $O(d^2)$ for arbitrary databases can be breached by exploiting the sparsity of the database. This motivates a more refined measure of error that takes the sparsity of A into account. Given an A and n , one can ask for the mechanism \mathcal{M} that minimizes the sparse case error $\max_{x: \|x\|_1 \leq n} \mathbb{E}[\|\mathcal{M}(x) - Ax\|_2^2]$. The next set of questions we study address this measure.

Question 3 *Given A and n , can we approximate the optimal sparse case error (ε, δ) -DP mechanism for A when restricted to databases of size at most n ?*

Question 4 *Given A and n , can we approximate the optimal sparse case error ε -DP mechanism for A when restricted to databases of size at most n ?*

¹Here and in the rest of the introduction, we suppress the dependence of the error on ε and δ .

The gap between the $\tilde{O}(dn^{\frac{4}{3}})$ error ε -DP mechanism of [BLR08] and the $\tilde{O}(dn)$ error (ε, δ) -DP mechanism of [HR10] leads us to ask:

Question 5 *Is there an ε -DP mechanism with error $\tilde{O}(dn)$ for databases of size at most n ?*

1.1 Results

In this work, we answer Questions 1-5 above. Denote by $B_1^N \triangleq \{x \in \mathbb{R}^N : \|x\|_1 \leq 1\}$ the N -dimensional ℓ_1 ball. Recall that for any query matrix $A \in \mathbb{R}^{d \times N}$ and any set $X \subseteq \mathbb{R}^N$, the (worst-case expected squared) error of \mathcal{M} is defined as

$$\text{err}_{\mathcal{M}}(A, X) \triangleq \max_{x \in X} \mathbb{E}[\|\mathcal{M}(x) - Ax\|_2^2].$$

In this paper, we are interested in both the case when $X = \mathbb{R}^N$, called the *dense* case, and when $X = nB_1^N$ for $n < d$, called the *sparse* case. We also write $\text{err}_{\mathcal{M}}(A) \triangleq \text{err}_{\mathcal{M}}(A, \mathbb{R}^N)$ and $\text{err}_{\mathcal{M}}(A, n) \triangleq \text{err}_{\mathcal{M}}(A, nB_1^N)$.

Our first result is a simple and efficient mechanism that for query matrix A gives an $O(\log^2 d)$ approximation to the optimal error.

Theorem 1 *Given a query matrix $A \in \mathbb{R}^{d \times N}$, there is an efficient (ε, δ) -DP mechanism \mathcal{M} and an efficiently computable lower bound L_A such that*

- $\text{err}_{\mathcal{M}}(A) \leq O(\log^2 d \log 1/\delta) \cdot L_A$, and
- for any (ε, δ) -DP mechanism \mathcal{M}' , $\text{err}_{\mathcal{M}'}(A, d) \geq L_A$.

We also show that the gap of $\Omega(\log(N/d))$ between ε -DP and (ε, δ) -DP mechanisms shown in [HT10] is essentially the worst possible, within $\text{polylog}(d)$ factor, for linear queries. More precisely, the lower bound on ε -DP mechanisms used in [HT10] is always within $O(\log(N/d) \text{polylog}(d))$ of the lower bound L_A computed by our algorithm above. Let \mathcal{M}^* denote the ε -DP generalized K -norm mechanism in [HT10].

Theorem 2 *For any (ε, δ) -DP mechanism \mathcal{M} , $\text{err}_{\mathcal{M}}(A) = \Omega(1/(\log^{O(1)}(d) \log(N/d))) \text{err}_{\mathcal{M}^*}(A)$.*

We next move to the sparse case. Here we give results analogous to the dense case with a slightly worse approximation ratio.

Theorem 3 *Given $A \in \mathbb{R}^{d \times N}$ and a bound n , there is an efficient (ε, δ) -DP mechanism \mathcal{M} and an efficiently computable lower bound $L_{A,n}$ such that*

- $\text{err}_{\mathcal{M}}(A, n) \leq O(\log^{3/2} d \cdot \sqrt{\log N \log 1/\delta} + \log^2 d \log 1/\delta) \cdot L_{A,n}$, and
- For any (ε, δ) -DP mechanism \mathcal{M}' , $\text{err}_{\mathcal{M}'}(A, n) \geq L_{A,n}$.

Theorem 4 *Given $A \in \mathbb{R}^{d \times N}$ and a bound n , there is an efficient ε -DP mechanism \mathcal{M} and an efficiently computable lower bound $L_{A,n}$ such that*

- $\text{err}_{\mathcal{M}}(A, n) \leq O(\log^{O(1)} d \cdot \log^{3/2} N) \cdot L_{A,n}$, and
- For any ε -DP mechanism \mathcal{M}' , $\text{err}_{\mathcal{M}'}(A, n) \geq L_{A,n}$.

We remark that in these theorems, our upper bounds hold for all x with $\|x\|_1 \leq n$, whereas the lower bounds hold even when x is an integer vector.

The (ε, δ) -DP mechanism of Theorem 3 when run on any counting query has error no larger than the best known bounds [GRU12] for counting queries, up to constants (not ignoring logarithmic factors). The ε -DP mechanism of Theorem 4 when run on any counting query can be shown to have nearly the same asymptotics, answering question 5 in the affirmative.

Theorem 5 *For any counting query A , there is an ε -DP mechanism \mathcal{M} such that $\text{err}_{\mathcal{M}}(A, n) = \tilde{O}(dn)$.*

We will summarize some key ideas we use to achieve these results. More details will follow in Section 1.2.

For the upper bounds, the first crucial step is to decompose A into “geometrically nice” components and then add Gaussian noise to each component. This is similar to the approach in [HT10,BDKT12] but we use the minimum volume enclosing ellipsoid, rather than the M -ellipsoid used in those works, to facilitate the decomposition process. This allows us to handle the approximate and the sparse cases. In addition, it simplifies the mechanism as well as the analysis. For the sparse case, we further couple the mechanism with least squares estimation of the noisy answer with respect to nAB_1^N . By utilizing techniques from statistical estimation, we can show that this process can reduce the error when $n < d$, and prove an error upper bound dependent on the size of the smallest projection of nAB_1^N .

For the lower bounds, we first lower bound the accuracy of (ϵ, δ) -DP mechanism by the hereditary discrepancy of the query matrix A , which we in turn lower bound in terms of the least singular values of submatrices of A . Finally, we close the loop by utilizing the restricted invertibility principle by Bourgain and Tzafriri [BT87] and its extension by Vershynin [Ver01] which, informally, shows that if there does not exist a “small” projection of nAB_1^N then A has a “large” submatrix with a “large” least singular value.

Approximating Hereditary Discrepancy

The discrepancy of a matrix $A \in \mathbb{R}^{d \times N}$ is defined to be $\text{disc}(A) = \min_{x \in \{-1, +1\}^N} \|Ax\|_\infty$. The hereditary discrepancy of a matrix is defined as $\text{herdisc}(A) = \max_{S \subseteq [N]} \text{disc}(A|_S)$, where $A|_S$ denotes the matrix A restricted to the columns indexed by S .

As hereditary discrepancy is a maximum over exponentially many submatrices, it is not a priori clear if there even exists a polynomial-time verifiable certificate for low hereditary discrepancy. Additionally, we can show that it is NP-hard to approximate hereditary discrepancy to within a factor of $3/2$. Bansal [Ban10] gave a pseudo-approximation algorithm for hereditary discrepancy, which efficiently computes a coloring of discrepancy at most a factor of $O(\log dN)$ larger than $\text{herdisc}(A)$ for a $d \times N$ matrix A . His algorithm allows efficiently computing a lower bound on herdisc for any restriction $A|_S$; however, such a lower bound may be arbitrarily loose, and before our work it was not known how to efficiently compute nearly matching lower and upper bounds on herdisc .

Muthukrishnan and Nikolov [MN12] show that for a query matrix $A \in \mathbb{R}^{d \times N}$, the error of any (ϵ, δ) -DP mechanism is lower bounded by (an ℓ_2^2 version of) $(\text{herdisc}(A))^2$ (up to logarithmic factors). Moreover, the lower bound used in Theorem 1 is in fact a lower bound on this version of $\text{herdisc}(A)$. Using the von Neumann minimax theorem, we can go between the ℓ_2^2 and the ℓ_∞ versions of these concepts, allowing us to sandwich the hereditary discrepancy of A between two quantities: a determinant based lower bound and the efficiently computable expected error of the private mechanism. As the two quantities are nearly matching, our work therefore leads to a polylogarithmic approximation to the hereditary discrepancy of any matrix A .

1.2 Techniques

In addition to known techniques from the differential privacy literature, our work borrows tools from discrepancy theory, convex geometry and statistical estimation. We next briefly describe how they fit in.

Central to designing a provably good approximation algorithm is an efficiently computable lower bound on the optimum. Muthukrishnan and Nikolov [MN12] proved that (a slight variant of) the hereditary discrepancy of A leads to a lower bound for the error of any (ϵ, δ) -DP mechanism. Lovász, Spencer and Vesztergombi [LSV86] showed that hereditary discrepancy itself can be lower bounded by a quantity called the determinant lower bound. Geometrically, this lower bound corresponds to picking the d columns of A that (along with the origin) give us a simplex with the largest possible volume. The volume or this simplex, appropriately normalized, gives us a lower bound on OPT. More precisely for any simplex S , $d^3 \cdot \text{vol}(S)^{\frac{2}{d}} \log^2 d$ gives a lower bound on the error. The $\log^2 d$ factor can be removed by using a lower bound based on the least singular values of submatrices of A . Geometrically, for the least singular value lower bound we need to find a simplex of large volume whose d non-zero vertices are also nearly pairwise orthogonal.

If the N columns of A all lie in a unit ball of radius R , it can be shown that adding Gaussian noise proportional to R suffices to guarantee (ϵ, δ) -DP, resulting in a mechanism having total squared error dR^2 . Can we relate this quantity to the lower bound? It turns out that if the unit ball of radius R is the minimum volume ellipsoid containing the columns of A , this can be done. In this case, a result of Vershynin [Ver01], building on the restricted invertibility results by Bourgain and Tzafriri [BT87], tells us that one can find $\Omega(d)$ vertices of K that touch the minimum containing ellipsoid, and are nearly orthogonal. The simplex formed by these vertices therefore has large volume, giving us a (ϵ, δ) -DP lower bound of $\Omega(dR^2)$. In this case, the Gaussian mechanism with the

optimal R is within a constant factor of the lower bound. When the minimum volume enclosing ellipsoid is not a ball, we need to project the query along the $\frac{d}{2}$ shortest axes of this ellipsoid, answer this projection using the Gaussian mechanism, and recurse on the orthogonal projection. Using the full power of the restricted invertability result by Vershynin allows us to construct a large simplex and prove our competitive ratio.

Hardt and Talwar [HT10] also used a volume based lower bound, but for ε -DP mechanisms, one can take K , the symmetric convex hull of all the columns of A and use its volume instead of the volume of S in the lower bound above. How do these lower bounds compare? By a result of Bárány and Füredi [BF88] and Gluskin [Glu07], one can show that the volume of the convex hull of N points can be bounded by $(\log N)^{d/2} d^{-d/2}$ times that of the minimum enclosing ellipsoid. This, along with the aforementioned restricted invertability results, allows us to prove that the ε -DP lower bound is within $O((\log N) \text{polylog } d)$ of the (ε, δ) -DP lower bound.

How do we handle sparse queries? The first observation is that the lower bounding technique gives us d columns of A and the resulting lower bound holds not just for A but even for the $d \times d$ submatrix of A corresponding to the maximum volume simplex S ; moreover, the lower bound holds even when all databases are restricted to $O(d)$ individuals. Thus the lower bound holds when $n = O(d)$ and this value marks the transition between the sparse and the dense cases. Moreover, when the minimum volume ellipsoid containing the columns of A is a ball, the restricted invertability principle of Bourgain and Tzafriri and Vershynin gives us a d -dimensional simplex with nearly pairwise orthogonal vertices, and, therefore any n -dimensional face of this simplex is another simplex of large volume. The large n -dimensional simplex gives a lower bound on error when databases are restricted to have at most n individuals.

For smaller n , the error added by the Gaussian mechanism may be too large, and even though the value Ax lies in nAB_1^N , the noisy answer will likely fall outside this set. A common technique in statistical estimation for handling such error is to “project” the noisy point back into nAB_1^N , i.e. report the point \hat{y} in nAB_1^N that minimizes the Euclidean distance to the noisy answer \tilde{y} . This projection step provably reduces the expected error! Geometrically, we use well known techniques from statistics to show that the error after projection is bounded by the “shadow” that nAB_1^N leaves on the noise vector; this shadow is much smaller than the length of the noise vector when $n = o(d)$. In fact, when the noise is a spherical Gaussian, it can be shown that $\|\hat{y} - y\|_2^2$ is only about $\frac{n}{d} \|\tilde{y} - y\|_2^2$. This gives near optimal bounds for the case when the minimum volume ellipsoid is a ball; the general case is handled using a recursive mechanism as before.

To get an ε -DP mechanism, we use the K -norm mechanism [HT10] instead of Gaussian noise. To bound the shadow of nAB_1^N on w , where w is the noise vector generated by the K -norm mechanism, we first analyze the expectation of $\langle a_i, w \rangle$ for any column of A , and we use the log concavity of the noise distribution to prove concentration of this random variable. A union bound helps complete the argument as in the Gaussian case.

1.3 Related Work

Dwork et al. [DMNS06] showed that any query can be released while adding noise proportional to the total *sensitivity* of the query. This motivated the question of designing mechanisms with good guarantees for any set of low sensitivity queries. Nissim, Raskhodnikova and Smith [NRS07] showed that adding noise proportional to (a smoothed version of) the *local sensitivity* of the query suffices for guaranteeing differential privacy; this may be much smaller than the worst case sensitivity for non-linear queries. Lower bounds on the amount of noise needed for general low sensitivity queries have been shown in [DN03, DMT07, DY08, DMNS06, RHS07, HT10, De12]. Kasiviswathan et al. [KRSU10] showed upper and lower bounds for contingency table queries and more recently [KRS13] showed lower bounds on publishing error rates of classifiers or even M-estimators. Muthukrishnan and Nikolov [MN12] showed that combinatorial discrepancy lower bounds the noise for answering any set of linear queries.

Using learning theoretic techniques, Blum, Ligett and Roth [BLR08] first showed that one can exploit sparsity of the database, and answer a large number of counting queries with error small compared to the number of individuals in the database. This line of work has been further extended and improved in terms of error bounds, efficiency, generality and interactivity in several subsequent works [DNR⁺09, DRV10, RR10, HR10, GHRU11, HLM12].

Ghosh, Roughgarden and Sundarajan [GRS09] showed that for any one dimensional counting query, a discrete version of the Laplacian mechanism is optimal for pure privacy in a very general utilitarian framework and Gupte and Sundarajan [GS10] extended this to risk averse agents. Brenner and Nissim [BN10] showed that such universally optimal private mechanisms do not exist for two counting queries or for a single non-binary sum query. As mentioned above, Hardt and Talwar [HT10], and Bhaskara et al. [BDKT12] gave relative guarantees for multi-dimensional queries under pure privacy with respect to total squared error. De [De12] unified and

strengthened these bounds and showed stronger lower bounds for the class of non-linear low sensitivity queries.

For specific queries of interest, improved upper bounds are known. Barak et al. [BCD⁺07] studied low dimensional marginals and showed that by running the Laplace mechanism on a different set of queries, one can reduce error. Using a similar strategy, improved mechanisms were given by [XWG10, CSS10] for orthogonal counting queries, and near optimal mechanisms were given by Muthukrishnan and Nikolov [MN12] for halfspace counting queries. The approach of answering a set of queries different from the target query set has also been studied in more generality and for other sets of queries by [LHR⁺10, DWHL11, RHS07, XWG10, XXY10, YZW⁺12]. Li and Miklau [LM12a, LM12b] study a class of mechanisms called extended matrix mechanisms and show that one can efficiently find the best mechanisms from this class. Hay et al. [HRMS10] show that in certain settings such as unattributed histograms, correcting noisy answers to enforce a consistency constraint can improve accuracy.

Very recently, Fawaz et al. [FMN] used the hereditary discrepancy lower bounds of Muthukrishnan and Nikolov, as well as the determinant lower bound on discrepancy of Lovasz, Spencer, and Vesztergombi, to prove that a certain Gaussian noise mechanism is nearly optimal (in the dense setting) for computing any given convolution map. Like our algorithms, their algorithm adds correlated Gaussian noise; however, they always use the Fourier basis to correlate the noise.

We refer the reader to texts by Chazelle [Cha00] and Matoušek [Mat99] and the chapter by Beck and Sós [BS95] for an introduction to discrepancy theory. Bansal [Ban10] showed that a semidefinite relaxation can be used to design a pseudo-approximation algorithm for hereditary discrepancy. Matoušek [Mat11] showed that the determinant based lower bound of Lovász, Spencer and Vesztergombi [LSV86] is tight up to polylogarithmic factors. Larsen [Lar11] showed applications of hereditary discrepancy to data structure lower bounds, and Chandrasekaran and Vempala [CV11] recently showed applications of hereditary discrepancy to problems in integer programming.

Roadmap. In Section 2.3.3 we introduce relevant preliminaries. In Section 3 we present our main results for approximate differential privacy, and in Section 4 we present our main results for pure differential privacy. In Section 5 we prove absolute upper bounds on the error required for privately answering sets of d counting queries. In Section 6 we give some extensions and applications of our main results, namely an optimal efficient mechanism for ℓ_∞ error in the dense case, and the efficient approximation to hereditary discrepancy implied by that mechanism. We conclude in Section 7.

2 Preliminaries

We start by introducing some basic notation.

Let B_1^d , and B_2^d be, respectively, the ℓ_1 and ℓ_2 unit balls in \mathbb{R}^d . Also, let $\text{sym}\{a_1, \dots, a_N\}$ be the convex hull of the vectors $\pm a_1, \dots, \pm a_N$. Equivalently, $\text{sym}\{a_1, \dots, a_N\} = AB_1^N$ where A is a matrix whose columns equal a_1, \dots, a_N .

For a $d \times N$ matrix A and a set $S \subseteq [N]$, we denote by $A|_S$ the submatrix of A consisting of those columns of A indexed by elements of S . Occasionally we refer to a matrix V whose columns form an orthonormal basis for some subspace of interest \mathcal{V} as the orthonormal basis of \mathcal{V} . \mathcal{P}_k is the set of orthogonal projections onto k -dimensional subspaces of \mathbb{R}^d .

By $\sigma_{\min}(A)$ and $\sigma_{\max}(A)$ we denote, respectively, the smallest and largest singular value of A . I.e., $\sigma_{\min}(A) = \min_{x: \|x\|_2=1} \|Ax\|_2$ and $\sigma_{\max}(A) = \max_{x: \|x\|_2=1} \|Ax\|_2$. In general, $\sigma_i(A)$ is the i -th largest singular value of A , and $\lambda_i(A)$ is the i -th largest eigenvalue of A . We recall the minimax characterization of eigenvalues for symmetric matrices:

$$\lambda_i = \max_{\mathcal{V}: \dim \mathcal{V}=i} \min_{x \in \mathcal{V}: \|x\|_2=1} x^T A x.$$

For a matrix A (and the corresponding linear operator), we denote by $\|A\|_2 = \sigma_{\max}(A)$ the spectral norm of A and $\|A\|_F = \sqrt{\sum_i \sigma_i^2(A)} = \sqrt{\sum_{i,j} a_{i,j}^2}$ the Frobenius norm of A . By $\ker A$ we denote the kernel of A , i.e. the subspace of vectors x for which $Ax = 0$.

2.1 Geometry

For a set $K \subseteq \mathbb{R}^d$, we denote by $\text{vol}_d(K)$ its d -dimensional volume. Often we use instead the *volume radius*

$$\text{vrad}_d(K) \triangleq (\text{vol}(K) / \text{vol}(B_2^d))^{1/d}.$$

Subscripts are omitted when this does not cause confusion. When K lies in a k -dimensional affine subspace of \mathbb{R}^d , $\text{vol}(K)$ and $\text{vrad}(K)$ (without subscripts) are understood to imply vol_k and vrad_k , respectively.

For a convex body $K \subseteq \mathbb{R}^d$, the *polar body* K° is defined by $K^\circ = \{y : \langle y, x \rangle \leq 1 \ \forall x \in K\}$. The fundamental fact about polar bodies we use is that for any two convex bodies K and L

$$K \subseteq L \Leftrightarrow L^\circ \subseteq K^\circ. \quad (1)$$

In the remainder of this paper, when we claim that a fact follows “by convex duality,” we mean that it is implied by (1).

A convex body K is (*centrally*) *symmetric* if $-K = K$. The *Minkowski norm* $\|x\|_K$ induced by a symmetric convex body K is defined as $\|x\|_K \triangleq \min\{r \in \mathbb{R} : x \in rK\}$. The Minkowski norm induced by the polar body K° of K is the *dual norm* of $\|x\|_K$ and also has the form $\|y\|_{K^\circ} = \max_{x \in K} \langle x, y \rangle$. For convex symmetric K , the induced norm and dual norm satisfy Hölder’s inequality:

$$|\langle x, y \rangle| \leq \|x\|_K \|y\|_{K^\circ}. \quad (2)$$

An *ellipsoid* in \mathbb{R}^d is the image of B_2^d under an affine map. All ellipsoids we consider are symmetric, and therefore, are equal to an image FB_2^d of the ball B_2^d under a linear map F . A full dimensional ellipsoid $E = FB_2^d$ can be equivalently defined as $E = \{x : x^T(FF^T)^{-1}x \leq 1\}$. The polar body of a symmetric ellipsoid $E = FB_2^d$ is the ellipsoid (or cylinder with an ellipsoid as its base in case F is not full dimensional) $E^\circ = \{x : x^T FF^T x \leq 1\}$.

We repeatedly use a classical theorem of Fritz John, characterizing the (unique) *minimum volume enclosing ellipsoid* (MEE) of any convex body K . We note that John’s theorem is frequently stated in terms of the maximum volume enclosed ellipsoid in K ; the two variants of the theorem are equivalent by convex duality. The MEE of K is also known as a the Löwner or Löwner-John ellipsoid of K .

Theorem 6 ([Joh48]) *Any convex body $K \subseteq \mathbb{R}^d$ is contained in a unique ellipsoid of minimal volume. This ellipsoid is B_2^d if and only if there exist unit vectors $u_1, \dots, u_m \in K \cap B_2^d$ and positive reals c_1, \dots, c_m such that*

$$\begin{aligned} \sum c_i u_i &= 0 \\ \sum c_i u_i u_i^T &= I \end{aligned}$$

According to John’s characterization, when the MEE of K is the ball B_2^d , the contact points of K and B_2^d satisfy a structural property — the identity decomposes into a linear combination of the projection matrices onto the lines of the contact points. Intuitively, this means that K “hits” B_2^d in all directions — it has to, or otherwise B_2^d can be “pinched” in order to produce a smaller ellipsoid that still contains K . This intuition is formalized by a theorem of Vershynin, which generalizes the work of Bourgain and Tzafriri on restricted invertibility [BT87]. Vershynin ([Ver01] Theorem 3.1) shows that there exist $\Omega(d)$ contact points of K and B_2^d which are approximately pairwise orthogonal.

Theorem 7 ([Ver01]) *Let $K \subseteq \mathbb{R}^d$ be a symmetric convex body whose minimum volume enclosing ellipsoid is the unit ball B_2^d . Let T be a linear map with spectral norm $\|T\|_2 \leq 1$. Then for any β , there exist constant $C_1(\beta)$, $C_2(\beta)$ and contact points x_1, \dots, x_k with $k \geq (1 - \beta)\|T\|_F^2$ such that the matrix $TX = (Tx_i)_{i=1}^k$ satisfies*

$$C_1(\beta) \frac{\|T\|_F}{\sqrt{d}} \leq \sigma_{\min}(TX) \leq \sigma_{\max}(TX) \leq C_2(\beta) \frac{\|T\|_F}{\sqrt{d}}$$

2.2 Statistical Estimation

A key element in our algorithms for the sparse case is the use of least squares estimation to reduce error. Below we present a bound on the error of least squares estimation with respect to symmetric convex bodies. This analysis appears to be standard in the statistics literature; a special case of it appears for example in [RWY11].

Lemma 1 *Let $L \subseteq \mathbb{R}^d$ be a symmetric convex body, and let $y \in L$ and $\tilde{y} = y + w$ for some $w \in \mathbb{R}^d$. Let, finally, $\hat{y} = \arg \min_{\hat{y} \in L} \|\hat{y} - \tilde{y}\|_2^2$. We have $\|\hat{y} - y\|_2^2 \leq \min\{4\|w\|_2^2, 4\|w\|_{L^\circ}\}$.*

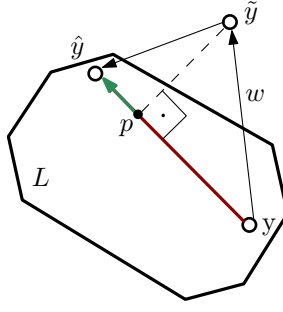


Figure 1: A schematic illustration of the proof of Lemma 1. The vector $p - y$ is proportional in length to $\langle \hat{y} - y, w \rangle$ and the vector $\hat{y} - p$ is proportional in length to $\langle \hat{y} - y, \hat{y} - \tilde{y} \rangle$. As $\|w\| \geq \|\hat{y} - \tilde{y}\|$, $\|p - y\| \geq \|\hat{y} - p\|$.

Proof: First we show the easier bound $\|\hat{y} - y\|_2 \leq 2\|w\|_2$, which follows by the triangle inequality:

$$\|\hat{y} - y\|_2 \leq \|\hat{y} - \tilde{y}\|_2 + \|\tilde{y} - y\|_2 \leq 2\|\tilde{y} - y\|_2.$$

The second bound is based on Hölder's inequality and the following simple but very useful fact, illustrated schematically in Figure 1:

$$\begin{aligned} \|\hat{y} - y\|_2^2 &= \langle \hat{y} - y, \tilde{y} - y \rangle + \langle \hat{y} - y, \hat{y} - \tilde{y} \rangle \\ &\leq 2\langle \hat{y} - y, \tilde{y} - y \rangle. \end{aligned} \quad (3)$$

The inequality (3) follows from

$$\langle \hat{y} - y, \tilde{y} - y \rangle = \|\tilde{y} - y\|_2^2 + \langle \hat{y} - \tilde{y}, \tilde{y} - y \rangle \geq \|\hat{y} - \tilde{y}\|_2^2 + \langle \hat{y} - \tilde{y}, \tilde{y} - y \rangle = \langle \hat{y} - \tilde{y}, \hat{y} - y \rangle.$$

Inequality (3), $w = \tilde{y} - y$, and Hölder's inequality imply

$$\|\hat{y} - y\|_2^2 \leq 2\langle \hat{y} - y, w \rangle \leq 2\|\hat{y} - y\|_L \|w\|_{L^\circ} \leq 4\|w\|_{L^\circ},$$

which completes the proof. ■

2.3 Differential Privacy

Following recent work in differential privacy, we model private data as a database D of n rows, where each row of D contains information about an individual. Formally, a database D is a multiset of size n of elements of the universe $U = \{t_1, \dots, t_N\}$ of possible user types. Our algorithms take as input a histogram $x \in \mathbb{R}^N$ of the database D , where the i -th component x_i of x encodes the number of individuals in D of type t_i . Notice that in this histogram representation, we have $\|x\|_1 = n$ when D is a database of size n . Also, two neighboring databases D and D' that differ in the presence or absence of a single individual correspond to two histograms x and x' satisfying $\|x - x'\|_1 = 1$.

Through most of this paper, we work under the notion of *approximate differential privacy*. The definition follows.

Definition 1 ([DMNS06, DKM⁺06b]) *A (randomized) algorithm \mathcal{M} with input domain \mathbb{R}^N and output range Y is (ϵ, δ) -differentially private if for every n , every x, x' with $\|x - x'\|_1 \leq 1$, and every measurable $S \subseteq Y$, \mathcal{M} satisfies*

$$\Pr[\mathcal{M}(x) \in S] \leq e^\epsilon \Pr[\mathcal{M}(x') \in S] + \delta.$$

When $\delta = 0$, we are in the regime of *pure differential privacy*.

An important basic property of differential privacy is that the privacy guarantees degrade smoothly under composition and are not affected by post-processing.

Lemma 2 ([DMNS06, DKM⁺06b]) *Let \mathcal{M}_1 and \mathcal{M}_2 satisfy (ϵ_1, δ_1) - and (ϵ_2, δ_2) -differential privacy, respectively. Then the algorithm which on input \mathbf{x} outputs the tuple $(\mathcal{M}_1(\mathbf{x}), \mathcal{M}_2(\mathcal{M}_1(\mathbf{x}), \mathbf{x}))$ satisfies $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$ -differential privacy.*

2.3.1 Optimality for Linear Queries

In this paper we study the necessary and sufficient error incurred by differentially private algorithms for approximating *linear queries*. A set of d linear queries is given by a $d \times N$ *query matrix* or *workload* A ; the exact answers to the queries on a histogram x are given by the d -dimensional vector $y = Ax$.

We define error as total squared error. More precisely, for an algorithm \mathcal{M} and a subset $X \subseteq \mathbb{R}^N$, we define

$$\text{err}_{\mathcal{M}}(A, X) \triangleq \sup_{x \in X} \mathbb{E} \|Ax - \mathcal{M}(A, x)\|_2^2.$$

We also write $\text{err}_{\mathcal{M}}(A, nB_1^N)$ as $\text{err}_{\mathcal{M}}(A, n)$. The optimal error achievable by any (ε, δ) -differentially private algorithm for queries A and databases of size up to n is

$$\text{opt}_{\varepsilon, \delta}(A, n) \triangleq \inf_{\mathcal{M}} \text{err}_{\mathcal{M}}(A, n),$$

where the infimum is taken over all (ε, δ) -differentially private algorithms. When no restrictions are placed on the size n of the database, the appropriate notion of optimal error is $\text{opt}_{\varepsilon, \delta}(A) \triangleq \sup_n \text{opt}_{\varepsilon, \delta}(A, n)$. Similarly, for an algorithm \mathcal{M} , the error when database size is not bounded is $\text{err}_{\mathcal{M}}(A) \triangleq \sup_n \text{err}_{\mathcal{M}}(A, n)$. A priori it is not clear that these quantities are necessarily finite, but we will show that this is the case.

In order to get tight dependence on the privacy parameter ε in our analyses, we will use the following relationship between $\text{opt}_{\varepsilon, \delta}(A, n)$ and $\text{opt}_{\varepsilon', \delta'}(A, n)$.

Lemma 3 *For any ε , any $\delta < 1$, any integer k and for $\delta' \geq \frac{e^{k\varepsilon} - 1}{e^\varepsilon - 1} \delta$,*

$$\text{opt}_{\varepsilon, \delta}(A, n) \geq k^2 \text{opt}_{k\varepsilon, \delta'}(A, n/k).$$

Proof: Let \mathcal{M} be an (ε, δ) -differentially private algorithm achieving $\text{opt}_{\varepsilon, \delta}(A, n)$. We will use \mathcal{M} as a black box to construct a $(k\varepsilon, \delta')$ -differentially private algorithm \mathcal{M}' which satisfies the error guarantee $\text{err}_{\mathcal{M}'}(A, n/k) \leq \frac{1}{k^2} \text{err}_{\mathcal{M}}(A, n)$.

The algorithm \mathcal{M}' on input x satisfying $\|x\|_1 \leq n/k$ outputs $\frac{1}{k} \mathcal{M}(kx)$. We need to show that \mathcal{M}' satisfies $(k\varepsilon, \delta')$ -differential privacy. Let x and x' be two neighboring inputs to \mathcal{M}' , i.e. $\|x - x'\|_1 \leq 1$, and let S be a measurable subset of the output \mathcal{M}' . Denote $p_1 = \Pr[\mathcal{M}'(x) \in S]$ and $p_2 = \Pr[\mathcal{M}'(x') \in S]$. We need to show that $p_1 \leq e^{k\varepsilon} p_2 + \delta'$. To that end, define $x_0 = kx$, $x_1 = kx + (x' - x)$, $x_2 = kx + 2(x' - x)$, \dots , $x_k = kx'$. Applying the (ε, δ) -privacy guarantee of \mathcal{M} to each of the pairs of neighboring inputs $x_0, x_1, x_1, x_2, \dots, x_{k-1}, x_k$ in sequence gives us

$$p_1 \leq e^{k\varepsilon} p_2 + (1 + e^\varepsilon + \dots + e^{(k-1)\varepsilon}) \delta = e^{k\varepsilon} p_2 + \frac{e^{k\varepsilon} - 1}{e^\varepsilon - 1} \delta.$$

This finishes the proof of privacy for \mathcal{M}' . It is straightforward to verify that $\text{err}_{\mathcal{M}'}(A, n/k) \leq \frac{1}{k^2} \text{err}_{\mathcal{M}}(A, n)$. ■

Above, we state the error and optimal error definitions for histograms x , which can be arbitrary real vectors. All our algorithms work in this general setting. Recall, however, that the histograms arising from our definition of databases are integer vectors. Our lower bounds do hold against integer histograms as well. Therefore, defining err and opt in terms of integer histograms (i.e. taking $\text{err}_{\mathcal{M}}(A, n) \triangleq \text{err}_{\mathcal{M}}(A, nB_1^N \cap \mathbb{N}^N)$) does not change the asymptotics of our theorems.

2.3.2 Gaussian Noise Mechanism

A basic mechanism for achieving (ε, δ) -differential privacy for linear queries is adding appropriately scaled independent Gaussian noise to each query. This approach goes back to the work of Blum et al. [BDMN05], predating the definition of differential privacy. Next we define this basic mechanism formally and give a privacy guarantee. The privacy analysis of the Gaussian mechanism in the context of (ε, δ) -differential privacy was first given in [DKM⁺06b]. We give the full proof here for completeness.

Lemma 4 *Let $A = (a_i)_{i=1}^N$ be a $d \times N$ matrix such that $\forall i : \|a_i\|_2 \leq \sigma$. Then a mechanism which on input $x \in \mathbb{R}^N$ outputs $Ax + w$, where $w \sim N(0, \sigma \frac{1 + \sqrt{2 \ln(1/\delta)}}{\varepsilon})^d$, satisfies (ε, δ) -differential privacy.*

Proof: Let $C = \frac{1+\sqrt{2\ln(1/\delta)}}{\varepsilon}$ and let p be the probability density function of $N(0, C\sigma)^d$. Let also $K = AB_1$, so $\|x - x'\|_1 \in B_1$ implies $A(x - x') \in K \subseteq B_2^d$. Define

$$D_v(w) \triangleq \ln \frac{p(w)}{p(w+v)}.$$

We will prove that when $w \sim N(0, C\sigma)$, for all $v \in K$, $\Pr[|D_v(w)| > \varepsilon] \leq \delta$. This suffices to prove (ε, δ) -differential privacy. Indeed, let the algorithm output $Ax + w$ and fix any x' s.t. $\|x - x'\|_1 \leq 1$. Let $v = A(x - x') \in K$ and $S = \{w : |D_v(w)| > \varepsilon\}$. For any measurable $T \subseteq \mathbb{R}^d$ we have

$$\begin{aligned} \Pr[Ax + w \in T] &= \Pr[w \in T - Ax] \\ &= \int_{S \cap (T - Ax)} p(w) dw + \int_{\bar{S} \cap (T - Ax)} p(w) dw \\ &\leq \delta + e^\varepsilon \int_{\bar{S} \cap (T - Ax')} p(w) dw \\ &= \delta + e^\varepsilon \Pr[w \in T - Ax'] = \delta + e^\varepsilon \Pr[Ax' + w \in T]. \end{aligned}$$

We fix an arbitrary $v \in K$ and proceed to prove $|D_v(w)| \leq \varepsilon$ with probability at least $1 - \delta$. We will first compute $\mathbb{E}D_v(w)$ and then apply a tail bound. Recall that $p(w) \propto \exp(-\frac{1}{2C^2\sigma^2}\|w\|_2^2)$. Notice also that, since $v \in K$ can be written as $\sum_{i=1}^N \alpha_i a_i$ where $\sum |\alpha_i| \leq 1$, we have $\|v\|_2 \leq \sigma$. Then we can write

$$\begin{aligned} \mathbb{E}D_v(w) &= \mathbb{E} \frac{\|v + w\|_2^2 - \|w\|_2^2}{2C^2\sigma^2} \\ &= \mathbb{E} \frac{\|v\|^2 + 2v^T w}{2C^2\sigma^2} \leq \frac{1}{2C^2} \end{aligned}$$

Note that to bound $|D_v(w)|$ we simply need to bound $\frac{1}{C^2\sigma^2}v^T w$ from above and below. Since $\frac{1}{C^2\sigma^2}v^T w \sim N(0, \frac{\|v\|}{C\sigma})$, we can apply a Chernoff bound and we get

$$\Pr \left[|v^T w| > \frac{1}{C} \sqrt{2\ln(1/\delta)} \right] \leq \delta.$$

Therefore, with probability $1 - \delta$,

$$\frac{1/2C - \sqrt{2\ln(1/\delta)}}{C} \leq D_v(w) \leq \frac{1/2C + \sqrt{2\ln(1/\delta)}}{C}.$$

Substituting $C \geq \frac{1+\sqrt{2\ln(1/\delta)}}{\varepsilon}$ completes the proof. \blacksquare

The following corollary is a useful geometric generalization of Lemma 4.

Corollary 8 *Let $A = (a_i)_{i=1}^N$ be a $d \times N$ matrix of rank d and let $K = \text{sym}\{a_1, \dots, a_N\}$. Let $E = FB_2^d$ (F is a linear map) be an ellipsoid containing K . Then a mechanism that outputs $Ax + Fw$ where $w \sim N(0, \frac{1+\sqrt{2\ln(1/\delta)}}{\varepsilon})^d$ satisfies (ε, δ) -differential privacy.*

Proof: Since K is full dimensional (by $\text{rank} A = d$) and E contains K , E is full dimensional as well, and, therefore, F is an invertible linear map. Define $G = F^{-1}A$. For each column g_i of G , we have $\|g_i\|_2 \leq 1$. Therefore, by Lemma 4, a mechanism that outputs $Gx + w$ (where w is distributed as in the statement of the corollary) satisfies (ε, δ) -differential privacy. Therefore, $FGx + Fw = Ax + Fw$ is (ε, δ) -differentially private by the post-processing property of differential privacy. \blacksquare

We present a composition theorem, specific to composing Gaussian noise mechanisms. We note that a similar composition result in a much more general setting but with slightly inferior dependence on the parameters is proven in [DRV10].

Corollary 9 *Let $\mathcal{V}_1, \dots, \mathcal{V}_k$ be vector spaces of respective dimensions d_1, \dots, d_k , such that $\forall i \leq k-1, \mathcal{V}_{i+1} \subseteq \mathcal{V}_i^\perp$ and $d_1 + \dots + d_k = d$. Let $A = (a_i)_{i=1}^N$ be a $d \times N$ matrix of rank d and let $K = \text{sym}(a_1, \dots, a_N)$. Let Π_i be the projection matrix for \mathcal{V}_i and let $E_i = F_i B_2^{d_i} \subseteq \mathcal{V}_i$ be an ellipsoid such that $\Pi_i K \subseteq E_i$. Then the mechanism that outputs $Ax + \sqrt{k} \sum_{i=1}^k F_i w_i$ where for each i , $w_i \sim N(0, \frac{1+\sqrt{2\ln(1/\delta)}}{\varepsilon})^{d_i}$, satisfies (ε, δ) -differential privacy.*

Proof: Let $c(\varepsilon, \delta) = \frac{1 + \sqrt{2 \ln(1/\delta)}}{\varepsilon}$. Since the random variables $F_1 w_1, \dots, F_k w_k$ are pairwise independent Gaussian random variables, and $F_i w_i$ has covariance matrix $c(\varepsilon, \delta)^2 F_i F_i^T$, we have that $w = \sqrt{k} \sum_{i=1}^k F_i w_i$ is a Gaussian random variable with covariance $c(\varepsilon, \delta)^2 G$, where $G = k \sum_{i=1}^k F_i F_i^T$. By Corollary 8, it is sufficient to show that the ellipsoid $E = GB_2^d$ contains K . By convex duality, this is equivalent to showing $E^\circ \subseteq K^\circ$, which is in turn equivalent to $\forall x : \|x\|_{K^\circ} \leq \|x\|_{E^\circ}$. Recalling that $\|x\|_{E^\circ}^2 = x^T G G^T x$ and $\|x\|_{K^\circ} = \max_{y \in K} \langle y, x \rangle = \max_{j=1}^N \langle a_j, x \rangle$, we need to establish

$$\forall x \in \mathbb{R}^d, \forall j \in [N] : \langle a_j, x \rangle^2 \leq x^T G G^T x. \quad (4)$$

We proceed by establishing (4). Since for all i , $\Pi_i K \subseteq E_i$, by duality and the same reasoning as above, we have that for all i and j , $\langle \Pi_i a_j, x \rangle^2 \leq x^T F_i F_i^T x$. Therefore, by the Cauchy-Schwarz inequality,

$$\begin{aligned} \langle a_j, x \rangle^2 &= \left(\sum_{i=1}^k \langle \Pi_i a_j, x \rangle \right)^2 \\ &\leq k \sum_{i=1}^k \langle \Pi_i a_j, x \rangle^2 \\ &\leq k \sum_{i=1}^k x^T F_i F_i^T x = x^T G G^T x. \end{aligned}$$

This completes the proof. ■

2.3.3 Noise Lower Bounds

We will make extensive use of a lower bound on the noise complexity of (ε, δ) -differentially private mechanisms in terms of combinatorial discrepancy. First we need to define the notion of hereditary α -discrepancy:

$$\text{herdisc}_\alpha(A, n) \triangleq \max_{S \subseteq [N] : |S| \leq n} \min_{\substack{x \in \{-1, 0, +1\}^S \\ \|x\|_1 \geq \alpha |S|}} \|(A|_S)x\|_2.$$

We denote $\text{herdisc}(A) \triangleq \max_n \text{herdisc}_1(A, n)$. An equivalent notation is $\text{herdisc}^{\ell_2}(A)$. When the ℓ_2 norm is substituted with ℓ_∞ , we have the classical notion hereditary discrepancy, here denoted $\text{herdisc}^{\ell_\infty}(A)$.

Next we present the lower bound, which is a simple extension of the discrepancy lower bound on noise recently proved by Muthukrishnan and Nikolov [MN12].

Theorem 10 ([MN12]) *Let A be an $d \times N$ real matrix. For any constant α and sufficiently small constant $\varepsilon \leq \varepsilon(\alpha)$ and $\delta \leq \delta(\alpha)$,*

$$\text{opt}_{\varepsilon, \delta}(A, n) = \Omega(1) \text{herdisc}_\alpha(A, n)^2.$$

We further develop two lower bounds for $\text{herdisc}_\alpha(A, n)$ which are more convenient to work with. The first lower bound is by using spectral techniques. Observe first that, since the ℓ_2 -norm of any vector does not increase under projection, we have $\text{herdisc}_\alpha(A, n) \geq \text{herdisc}_\alpha(\Pi A, n)$ for any projection matrix Π . Furthermore, recall that for a matrix M , $\sigma_{\min}(M) = \min_{x: \|x\|_2=1} \|Mx\|_2$. For any $x \in \{-1, 0, 1\}^S$ satisfying $\|x\|_1 \geq \alpha |S|$, we have $\|x\|_2^2 \geq \alpha |S|$. Therefore,

$$\text{herdisc}_\alpha(A, n)^2 \geq \max_{S \subseteq [N] : |S| \leq n} \alpha |S| \sigma_{\min}^2(A|_S). \quad (5)$$

Let's define

$$\text{specLB}(A, n) \triangleq \max_{\substack{S \subseteq [N] \\ |S|=k \leq n}} \max_{\Pi \in \mathcal{P}_k} k \sigma_{\min}^2(\Pi A|_S).$$

Substituting (5) into Theorem 10, we have that there exist constants c_1 and c_2 such that

$$\text{opt}_{c_1, c_2}(A, n) = \Omega(1) \cdot \text{specLB}(A, n). \quad (6)$$

For the remainder of this paper we fix some constants c_1 and c_2 for which (6) holds. Similarly to the notation for opt , we will also sometimes denote $\text{specLB}(A) = \max_n \text{specLB}(A, n)$. We will use primarily the spectral lower bound (6) for arguing the optimality of our algorithms.

To show the small gap between the approximate and pure privacy (Theorem 2), we next develop a determinant based lower bound. We first switch from herdisc_α to the classical notion of hereditary discrepancy, equivalent to herdisc_1 , by observing the following relation between herdisc_α and herdisc_1 from [MN12]:

$$\text{herdisc}_1(A, n) \leq \frac{\log n}{\log \frac{1}{1-\alpha}} \text{herdisc}_\alpha(A, n) \quad (7)$$

We then use an extension of the classical determinant lower bound for hereditary discrepancy, due to Lovász, Spencer, and Vesztergombi.

Theorem 11 ([LSV86]) *For any real $d \times N$ matrix A ,*

$$(\text{herdisc}_1(A, n))^2 \geq \Omega(1) \cdot \det \text{LB}(A, n) \triangleq \max_{k \leq n} \max_{\substack{\Pi \in \mathcal{P}_k \\ S \subseteq [N]: |S|=k}} k \cdot |\det(\Pi A|_S)|^{2/k}.$$

Proof: The proof proceeds in two steps: showing that a quantity known as linear discrepancy is at most a constant factor larger than hereditary discrepancy; lower bounding linear discrepancy in terms of $\det \text{LB}$. Most of the proof can be adapted with little modification from proofs of the lower bound on ℓ_∞ discrepancy in [LSV86]. Here we follow the exposition in [Cha00].

Let us first define linear discrepancy. For a $d \times k$ matrix M and $c \in [-1, 1]^k$, let $\text{disc}^c(M)$ be defined as

$$\text{disc}^c(M) \triangleq \min_{x \in \{-1, 1\}^k} \|Mx - Mc\|_2$$

The linear discrepancy of M is then defined as $\text{lindisc}(M) \triangleq \max_{c \in [-1, 1]^k} \text{disc}^c(M)$. We claim that for any M ,

$$\text{lindisc}(M) \leq 2 \text{herdisc}(M). \quad (8)$$

The bound (8) is proven for the more common ℓ_∞ variants of lindisc and herdisc in [Cha00], but the proof can be seen to apply without modification to discrepancy defined with respect to any norm. For completeness, we give the full argument here. For $c \in \{-1, 0, 1\}^k$, we have $\text{disc}^c(M) \leq \text{herdisc}(M)$. Call a vector $c \in [-1, 1]^k$ q -integral if any coordinate c_i of c can be written as $\sum_{j=0}^q b_j 2^{-j}$ where $(b_j)_{j=0}^q \in \{0, 1\}^{q+1}$. In other words, c is q -integral if the binary expansion of any coordinate of c is identically zero after the q -th digit. The bound (8) holds for 0-integral c , and we prove that it holds for q -integral c by induction on q . For the induction step, assume that the bound holds for any $(q-1)$ -integral c' and let c be q -integral. Define $s \in \{-1, 1\}^n$ by setting $s_i = -1$ if $c_i \geq 0$ and $s_i = 1$ otherwise. Then $c' = 2c + s \in [-1, 1]^k$ is $(q-1)$ -integral, and, by the induction hypothesis, there exists $x_1 \in \{-1, 1\}^k$ such that $\|Mx_1 - Mc'\|_2 \leq 2 \text{herdisc}(M)$. Let $c'' = \frac{x_1 - s}{2} \in \{-1, 0, 1\}^k$. Dividing by 2 and rearranging, we have $\|Mc'' - Mc\|_2 \leq \text{herdisc}(M)$. The vector c'' is 0-integral, and therefore there exists some $x_2 \in \{-1, 1\}^k$ such that $\|Mx_2 - Mc''\|_2 \leq \text{herdisc}(M)$. By the triangle inequality, $\|Mx_2 - Mc\|_2 \leq 2 \text{herdisc}(M)$, and this completes the inductive step.

We complete the proof of the theorem by proving a lower bound on lindisc in terms of $\det \text{LB}$. We note that a similar lower bound can be proved for any variant of lindisc defined in terms of any norm. The exact lower bound will depend on the volume radius of the unit ball of the norm. Since the proof of (8) also works for any norm, we get a determinant lower bound for hereditary discrepancy defined in terms of any norm as well.

We show that for any $d \times k$ matrix M

$$\text{lindisc}(M) = \Omega(1) \max_{\Pi \in \mathcal{P}_k} \sqrt{k} |\det \Pi M|^{1/k}. \quad (9)$$

Letting k range over $[n]$, M range over all $d \times k$ submatrices of A , and applying the bounds (8) and (9) implies the theorem.

We proceed to prove (9). Note that if $\text{rank}(M) < k$, (9) is trivially true; therefore, we may assume that $\text{rank}(M) = k$. Note also that without loss of generality we can take Π to be the orthogonal projection onto the range of M , since this is the projection operator that maximizes $|\det \Pi M|$. Let E be the ellipsoid $E = \{x : \|Mx\|_2^2 \leq 1\}$. The inequality $\text{lindisc}(M) \leq D$ is equivalent to

$$[-1, 1]^k \subseteq \bigcup_{x \in \{-1, 1\}^k} D \cdot E + x. \quad (10)$$

Thus $2^k = \text{vol}([-1, 1]^d) \leq 2^k \text{vol}(D \cdot E)$, and therefore $D^k \geq \frac{1}{\text{vol}(E)}$. On the other hand, the volume of E is equal to

$$\text{vol}(E) = \frac{\text{vol}(B_2)}{|\det(M^T M)|^{1/2}} = \frac{\text{vol}(B_2)}{|\det(\Pi M)|}.$$

Applying the standard estimate $\text{vol}(B_2^k)^{1/k} = \Theta(k^{-1/2})$ completes the proof. \blacksquare

By the determinant lower bound, and (7), we get our determinant lower bound on the noise necessary for privacy. For some constant $c_1, c_2 > 0$,

$$\text{opt}_{c_1, c_2}(A, n) = \Omega\left(\frac{1}{\log^2 n}\right) \text{detLB}(A, n). \quad (11)$$

Finally, we recall the stronger volume lower bound against $(\varepsilon, 0)$ -differential privacy from [HT10, BDKT12]. This lower bound is nearly optimal for $(\varepsilon, 0)$ -differential privacy, but does not hold for (ε, δ) -differential privacy when δ is $2^{-o(d)}$.

Theorem 12 ([HT10, BDKT12]) *For any $d \times N$ real matrix $A = (a_i)_{i=1}^N$,*

$$\text{opt}_{\varepsilon, 0}(A, \frac{d}{\varepsilon}) \geq \text{volLB}(A, \varepsilon) \triangleq \max_{k=1}^d \max_{\Pi \in \mathcal{P}_k} \frac{k^2}{\varepsilon^2} \text{vrad}(\Pi K)^2, \quad (12)$$

where $K = \text{sym}\{a_i\}_{i=1}^d$.

Furthermore, there exists an efficient mechanism \mathcal{M}_K (the generalized K -norm mechanism) which is $(\varepsilon, 0)$ -differentially private and satisfies $\text{err}_{\mathcal{M}_K}(A) = O(\log^3 d) \text{volLB}(A, \varepsilon)$.

3 Algorithms for Approximate Privacy

In this section we present our main results: efficient nearly optimal algorithms for approximate privacy in the cases of dense databases ($n > d/\varepsilon$) and sparse databases ($n = o(d/\varepsilon)$). Both algorithms rely on recursively computing an orthonormal basis for \mathbb{R}^d , based on the minimum volume enclosing ellipsoid of the columns of the query matrix A . We first present the algorithm for computing this basis, together with a property essential for the analyses of the two algorithms presented next.

3.1 The Base Decomposition Algorithm

We first present an algorithm (Algorithm 1) that, given a matrix $A \in \mathbb{R}^{d \times N}$, computes a set of orthonormal matrices U_1, \dots, U_k , where $k \leq \lceil 1 + \log d \rceil$. For each $i \neq j$, $U_i^T U_j = 0$, and the union of the columns of U_1, \dots, U_k forms an orthonormal basis for \mathbb{R}^d . Thus, Algorithm 1 computes a basis for \mathbb{R}^d , and partitions (“decomposes”) it into $k = O(\log d)$ bases of mutually orthogonal subspaces. This set of bases also induces a decomposition of A into $A = A_1 + \dots + A_k$, where $A_i = U_i U_i^T A$. The base decomposition of Algorithm 1 is essential to both our dense case and sparse case algorithms. Intuitively, for both cases we can show that the error of a simple mechanism applied to A_i can be matched by an error lower bound for $A_{i+1} + \dots + A_k$. The error lower bounds are based on the spectral lower bound specLB on discrepancy; the geometric properties of the minimum enclosing ellipsoid of a convex body together with the restricted invertibility principle of Bourgain and Tzafriri are key in deriving the lower bounds.

The next lemma captures the technical property of the decomposition of Algorithm 1 that allows us to prove matching upper and lower error bounds for our dense and sparse case algorithms.

Lemma 5 *Let d_i be the dimension of the span of U_i . Furthermore, for $i < k$, let $W_i = \sum_{j>i} U_j$, and let $W_k = U_k$. For every $i \leq k$, there exists a set $S_i \subseteq [N]$, such that $|S_i| = \Omega(d_i)$ and $\sigma_{\min}^2(W_i W_i^T A|_{S_i}) = \Omega(1) \max_{j=1}^N \|U_i^T a_j\|_2^2$.*

Proof: Let us, for ease of notation, assume that d is a power of 2. We prove that there exists a set S , $|S| = \Omega(d)$, such that

$$\sigma_{\min}^2(VV^T A|_S) = \Omega(1) \max_{j \in N} \|U_1^T a_j\|_2^2. \quad (13)$$

This proves the lemma for $i = 1$ (substitute $W_i = V$ and $d_i = d/2$). Applying (13) inductively to $V^T A$ establishes the lemma for all $i < k$. In case $i = k$, observe that $d_k = 1$ and we only need to show that the matrix $U_k U_k^T A$ has a column with ℓ_2 norm at least $\max_{j \in N} \|U_k^T a_j\|_2$. This is trivially true since each column of $U_k U_k^T A$ has the same ℓ_2 norm as the corresponding column of $U_k^T A$.

Algorithm 1 BASE DECOMPOSITION

Input $A = (a_i)_{i=1}^N \in \mathbb{R}^{d \times N}$ ($\text{rank} A = d$);
Compute $E = FB_2^d$, the minimum volume enclosing ellipsoid of $K = AB_1$;
Let $(u_i)_{i=1}^d$ be the (left) singular vectors of F corresponding to singular values $\sigma_1 \geq \dots \geq \sigma_d$;
if $d = 1$ **then**
 Output $U_1 = u_1$.
else
 Let $U_1 = (u_i)_{i>d/2}$ and $V = (u_i)_{i \leq d/2}$;
 Recursively compute a base decomposition V_2, \dots, V_k of $V^T A$ ($k \leq \lceil 1 + \log d \rceil$ is the depth of the recursion);

 For each $i > 1$, let $U_i = VV_i$;
 Output $\{U_1, \dots, U_k\}$.
end if

By applying an appropriate unitary transformation to the columns of A , we may assume that the major axes of E are co-linear with the standard basis vectors of \mathbb{R}^d , and, therefore, F is a diagonal matrix with $F_{ii} = \sigma_i$. This transformation comes without loss of generality, since it applies a unitary transformation to the columns of A and V and does not affect the singular values of any matrix $VV^T A|_S$ for any $S \subseteq [N]$. For the rest of the proof we assume that F is diagonal.

Since F is diagonal, u_i is equal to e_i , the i -th standard basis vector. Therefore U_1 is diagonal and equal to the projection onto $e_{d/2+1}, \dots, e_d$, and V is also diagonal and equal to the projection onto $e_1, \dots, e_{d/2}$. Consider $L = F^{-1}K = F^{-1}AB_1^d$ (recall that we assumed that $\text{rank} A = d$ and therefore F is non-singular). Since the minimum enclosing ellipsoid of K is $E = FB_2^d$, we have that the minimum enclosing ellipsoid of L is B_2^d . Let $T = VV^T$ be the projection onto $e_1, \dots, e_{d/2}$. Then, by Theorem 7, and because $\|T\|_F^2 = d/2$, we have that there exists a set S of size $|S| = \Omega(d)$ such that $\sigma_{\min}^2(TF^{-1}A|_S) = \Omega(1)$. We chose F , and therefore F^{-1} , as well as T to be diagonal matrices, so they all commute. Then, since T is a projection matrix,

$$\begin{aligned} \sigma_{\min}^2(VV^T A|_S) &= \sigma_{\min}^2(TA|_S) = \sigma_{\min}^2(FTF^{-1}A|_S) = \sigma_{\min}^2(FTTF^{-1}A|_S) \\ &\geq \sigma_{\min}^2(FT)\sigma_{\min}^2(TF^{-1}A|_S) = \Omega(\sigma_{d/2}^2). \end{aligned} \tag{14}$$

Observe that, since $K \subseteq E$, we have

$$U_1^T K \subseteq U_1^T E = U_1^T FB_2^d \subseteq \sigma_{d/2+1} B_2^d.$$

Therefore, $\max_{j=1}^N \|U_1^T a_j\|_2^2 \leq \sigma_{d/2+1}^2 \leq \sigma_{d/2}^2$. Substituting for $\sigma_{d/2}^2$ into (14) completes the proof. \blacksquare

3.2 The Dense Case: Correlated Gaussian Noise

Our first result is an efficient algorithm whose expected error matches the spectral lower bound specLB up to polylogarithmic factors and is therefore nearly optimal. This proves Theorem 1. The algorithm adds correlated unbiased Gaussian noise to the exact answer Ax . The noise distribution is computed based on the decomposition algorithm of the previous subsection.

Algorithm 2 GAUSSIAN NOISE MECHANISM

Input (*Public*): query matrix $A = (a_i)_{i=1}^N \in \mathbb{R}^{d \times N}$ ($\text{rank} A = d$);

Input (*Private*): database $x \in \mathbb{R}^N$

Let U_1, \dots, U_k be base decomposition computed by Algorithm 1 on input A , where U_i is an orthonormal basis for a space of dimension d_i ;

Let $c(\varepsilon, \delta) = \frac{1 + \sqrt{2 \ln(1/\delta)}}{\varepsilon}$.

For each i , let $r_i = \max_{j=1}^N \|U_i^T a_j\|_2$;

For each i , sample $w_i \sim N(0, c(\varepsilon, \delta))^{d_i}$;

Output $Ax + \sqrt{k} \sum_{i=1}^k r_i U_i w_i$.

Theorem 13 Let $\mathcal{M}_g(A, x)$ be the output of Algorithm 2 on input a $d \times N$ query matrix A and private input x . $\mathcal{M}_g(A, x)$ is (ε, δ) -differentially private and for all small enough ε and all δ small enough with respect to ε satisfies

$$\text{err}_{\mathcal{M}_g}(A) = O(\log^2 d \log 1/\delta) \text{opt}_{\varepsilon, \delta}(A, \frac{d}{\varepsilon}) = O(\log^2 d \log 1/\delta) \text{opt}_{\varepsilon, \delta}(A).$$

Notice that even though we assume $\text{rank} A = d$, this is without loss of generality: if $\text{rank} A = r < d$, we can compute an orthonormal basis V for the range of A and apply the algorithm to $A' = V^T A$ to compute an approximation \tilde{z} to $A'x$. We have that $VA'x = VV^T Ax = Ax$ since VV^T is a projection onto the range of A and Ax belongs to that range. Then $\tilde{y} = V\tilde{z}$ gives us an approximation of Ax satisfying $\|\tilde{y} - Ax\|_2 = \|\tilde{z} - A'x\|_2$.

We start the proof of Theorem 13 with the privacy analysis. For ease of notation, we assume throughout the analysis that d is a power of 2.

Lemma 6 $\mathcal{M}_g(A, x)$ satisfies (ε, δ) -differential privacy.

Proof: The lemma follows from Corollary 9. Next we describe in detail why the corollary applies.

Let U_1, \dots, U_k be the base decomposition computed by Algorithm 1 on input A . Let \mathcal{V}_i be the subspace spanned by the columns of U_i and let d_i be the dimension of \mathcal{V}_i . The projection matrix onto \mathcal{V}_i is $U_i U_i^T$. Let E_i be the ellipsoid $U_i(r_i B_2^{d_i}) = F_i B_2^{d_i}$ (F_i is $r_i U_i$). By the definition of r_i , $U_i^T K \subseteq r_i B_2^{d_i}$, and therefore, $\Pi_i K \subseteq E_i$. $\mathcal{M}_g(A, x)$ is distributed identically to $Ax + \sqrt{k} \sum_{i=1}^k F_i w_i$. Therefore, by Corollary 9, $\mathcal{M}_g(A, x)$ satisfies (ε, δ) -differential privacy. \blacksquare

Lemma 7 For all small enough ε and all δ small enough with respect to ε , for all i ,

$$\mathbb{E}\|r_i U_i w_i\|_2^2 \leq O(\log \frac{1}{\delta}) \text{opt}_{\varepsilon, \delta}(A, \frac{d_i}{\varepsilon}),$$

where r_i , U_i and w_i are as defined in Algorithm 2.

Proof:

To upper bound $\mathbb{E}\|r_i U_i w_i\|_2^2$, observe that, since the columns of U_i are d_i pairwise orthogonal unit vectors, $\|r_i U_i w_i\|_2^2 = r_i^2 \|w_i\|_2^2$. Therefore, it follows that

$$\mathbb{E}\|r_i U_i w_i\|_2^2 = d_i c(\varepsilon, \delta)^2 r_i^2 = O(\log(1/\delta) \frac{1}{\varepsilon^2} d r_i^2). \quad (15)$$

By (15), it is enough to lower bound $\text{opt}_{\varepsilon, \delta}(A)$ by $\Omega(\frac{1}{\varepsilon^2} d r_i^2)$. As a first step we lower bound $\text{specLB}(A)$. Then the lower bound on $\text{opt}_{\varepsilon, \delta}$ will follow from (6) and Lemma 3.

To lower bound $\text{specLB}(A)$, we invoke Lemma 5. It follows from the lemma that for every i there exists a projection matrix $\Pi_i = W_i W_i^T$ and a set S_i such that $\sigma_{\min}^2(\Pi_i A|_{S_i}) = \Omega(r_i^2)$, and, furthermore, $|S_i| = \Omega(d_i)$. Substituting into the definition of $\text{specLB}(A, d_i)$, we have that for all i .

$$\text{specLB}(A, d_i) \geq |S_i| \sigma_{\min}^2(\Pi_i A|_{S_i}) = \Omega(d_i r_i^2).$$

Therefore, by (6), $\text{opt}_{c_1, c_2}(A, d_i) = \Omega(d_i r_i^2)$ for all i . Finally, by Lemma 3, there exists a small enough $\delta = \delta(\varepsilon)$, for which $\text{opt}_{\varepsilon, \delta}(A, \frac{d_i}{\varepsilon}) = \Omega(\frac{1}{\varepsilon^2} d_i r_i^2)$, and this completes the proof. \blacksquare

Proof of Theorem 13: The proof of the theorem follows from Lemma 6 and Lemma 7. The privacy guarantee is direct from Lemma 6. Next we prove that the error of \mathcal{M}_g is near optimal.

Let w be the noise vector generated by Algorithm 2 so that $w = \sqrt{k} \sum_{i=1}^{1+\log d} r_i U_i w_i$. We have $\text{err}_{\mathcal{M}_g}(A) = \mathbb{E}\|w\|_2^2$. We proceed to upper bound this quantity in terms of $\text{opt}_{\varepsilon, \delta}(A)$.

By Lemma 7, for each w_i , $\mathbb{E}\|r_i U_i w_i\|_2^2 = O(\log 1/\delta) \text{opt}_{\varepsilon, \delta}(A)$. Since $U_i^T U_j = 0$ for all $i \neq j$, and $k = O(\log d)$, we can bound $\mathbb{E}\|w\|_2^2$ as

$$\mathbb{E}\|w\|_2^2 = k \sum_{i=1}^{1+\log d} \mathbb{E}\|r_i U_i w_i\|_2^2 = O(\log d \log 1/\delta) \sum_{i=1}^k \text{opt}_{\varepsilon, \delta}(A, \frac{d_i}{\varepsilon}) = O(\log^2 d \log 1/\delta) \text{opt}_{\varepsilon, \delta}(A, \frac{d}{\varepsilon})$$

This completes the proof. \blacksquare

3.3 The Sparse Case: Least Squares Estimation

In this subsection we present an algorithm with stronger accuracy guarantees than Algorithm 2: it is optimal for any query matrix A and any database size bound n (Theorem 3). The algorithm combines the noise distribution of Algorithm 2 with a least squares estimation step. Privacy is guaranteed by noise addition, while the least squares estimation step reduces the error significantly when $n = o(d/\varepsilon)$. The algorithm is shown as Algorithm 3.

Algorithm 3 LEAST SQUARES MECHANISM

Input (*Public*): query matrix $A = (a_i)_{i=1}^N \in \mathbb{R}^{d \times N}$ ($\text{rank} A = d$); database size bound n

Input (*Private*): database $x \in \mathbb{R}^N$

Let U_1, \dots, U_k be base decomposition computed by Algorithm 1 on input A , where U_i is an orthonormal basis for a space of dimension d_i ;

Let t be the largest integer such that $d_t \geq \varepsilon n$;

Let $X = \sum_{i=1}^t U_i$ and $Y = \sum_{i=t+1}^k U_i$;

Call Algorithm 2 to compute $\tilde{y} = \mathcal{M}_g(A, x)$;

Let $\tilde{y}_1 = XX^T \tilde{y}$ and $\tilde{y}_2 = YY^T \tilde{y}$;

Let $\hat{y}_1 = \arg \min \{\|\tilde{y}_1 - \hat{y}_1\|_2^2 : \hat{y}_1 \in nXX^TK\}$, where $K = AB_1$.

Output $\hat{y}_1 + \tilde{y}_2$.

Theorem 14 Let $\mathcal{M}_\ell(A, x, n)$ be the output of Algorithm 3 on input a $d \times N$ query matrix A , database size bound n and private input x . $\mathcal{M}_\ell(A, x)$ is (ε, δ) -differentially private and for all small enough ε and all δ small enough with respect to ε satisfies

$$\text{err}_{\mathcal{M}_\ell}(A) = O(\log^{3/2} d \sqrt{\log N \log(1/\delta)} + \log^2 d \log(1/\delta)) \text{opt}_{\varepsilon, \delta}(A, n).$$

Lemma 8 $\mathcal{M}_\ell(A, x, n)$ satisfies (ε, δ) -differential privacy.

Proof: The output of $\mathcal{M}_\ell(A, x, n)$ is a deterministic function of the output of $\mathcal{M}_g(A, x)$. By Lemma 6, $\mathcal{M}_g(A, x)$ satisfies (ε, δ) -differential privacy, and, therefore, by Lemma 2 (i.e. the post-processing property of differential privacy), $\mathcal{M}_\ell(A, x, n)$ satisfies (ε, δ) -differential privacy. \blacksquare

Lemma 9 Let U_i and t be as defined in Algorithm 3 and let r_i and w_i be as defined in Algorithm 2. Then $\mathbb{E} \max_{j=1}^N n |\langle a_j, r_i U_i w_i \rangle| \leq O(\sqrt{\log N \log(1/\delta)}) \text{opt}_{\varepsilon, \delta}(A, n)$.

Proof:

First we upper bound $\mathbb{E} \max_{j=1}^N n |\langle a_j, r_i U_i w_i \rangle|$. After rearranging the terms, we have $\langle a_j, r_i U_i w_i \rangle = \langle U_i^T a_j, r_i w_i \rangle$ for any i and j . By the definition of r_i , $\|U_i^T a_j\|_2 \leq r_i$. Therefore, $\langle U_i^T a_j, r_i w_i \rangle$ is a Gaussian random variable with mean 0 and standard deviation at most $r_i^2 c(\varepsilon, \delta)$. Using standard bounds on the expected maximum of N Gaussians (e.g. using a Chernoff bound and a union bound), we have that

$$\begin{aligned} \mathbb{E} \max_{j=1}^N n |\langle a_j, r_i U_i w_i \rangle| &= n \mathbb{E} \max_{j=1}^N |\langle U_i^T a_j, r_i w_i \rangle| \\ &= O(\sqrt{\log N} c(\varepsilon, \delta) n r_i^2) \\ &= O(\sqrt{\log N \log(1/\delta)} \frac{1}{\varepsilon} n r_i^2). \end{aligned}$$

To finish the proof of the lemma, we need to lower bound $\text{opt}_{\varepsilon, \delta}(A, n)$ by $\Omega(\frac{1}{\varepsilon} n r_i^2)$. We will use Lemma 5 to lower bound $\text{specLB}(A, \varepsilon n)$ by $\Omega(\varepsilon n r_i^2)$ and then we will invoke Lemma 3 to get the right dependence on ε .

By Lemma 5, for every i there exists a projection matrix $\Pi_i = W_i W_i^T$ and a set S_i such that $\sigma_{\min}^2(\Pi_i A|_{S_i}) = \Omega(r_i^2)$, and, furthermore, $|S_i| = \Omega(d_i)$. By the definition of t , for all $i \leq t$, $d_i \geq \varepsilon n$, and, therefore, $|S_i| = \Omega(\varepsilon n)$. Take $T_i \subseteq S_i$ to be an arbitrary subset of S_i of size $\Omega(\varepsilon n)$. The smallest singular value of $\Pi_i A|_{S_i}$ is a lower bound on the smallest singular value of $\Pi_i A|_{T_i}$:

$$\begin{aligned} \sigma_{\min}(\Pi_i A|_{T_i}) &= \min_{x: \|x\|_2=1} \|(\Pi_i A|_{T_i})x\|_2 = \min_{\substack{x: \|x\|_2=1 \\ \text{supp}(x) \subseteq T_i}} \|(\Pi_i A|_{S_i})x\|_2 \\ &\geq \min_{x: \|x\|_2=1} \|(\Pi_i A|_{S_i})x\|_2 = \sigma_{\min}(\Pi_i A|_{S_i}), \end{aligned}$$

where $\text{supp}(x)$ is the subset of coordinates on which x is nonzero. Therefore, $\sigma_{\min}^2(\Pi_i A|_{T_i}) = \Omega(r_i^2)$. Substituting into the definition of $\text{specLB}(A, \varepsilon n)$, we have

$$\text{specLB}(A, \varepsilon n) = \Omega(|T_i| \sigma_{\min}(\Pi_i A|_{T_i})) = \Omega(\varepsilon n r_i^2).$$

Therefore, by (6), $\text{opt}_{c_1, c_2}(A, \varepsilon n) = \Omega(\varepsilon n r_i^2)$ for all $i \leq t$. Finally, by Lemma 3, there exists a small enough $\delta = \delta(\varepsilon)$, for which $\text{opt}_{\varepsilon, \delta}(A, n) = \Omega(\frac{n}{\varepsilon} r_i^2)$, and this completes the proof. \blacksquare

Proof of Theorem 14: The privacy guarantee is direct from Lemma 8. Next we prove that the error of \mathcal{M}_ℓ is near optimal.

We will bound $\text{err}_{\mathcal{M}_\ell}(A, n)$ by bounding $\mathbb{E}\|\hat{y}_1 + \tilde{y}_2 - Ax\|_2^2$ for any x . Let us fix x and define $y = Ax$; furthermore, define $y_1 = XX^T y$ and $y_2 = YY^T y$. By the Pythagorean theorem, we can write

$$\mathbb{E}\|\hat{y}_1 + \tilde{y}_2 - y\|_2^2 = \mathbb{E}\|\hat{y}_1 - y_1\|_2^2 + \mathbb{E}\|\tilde{y}_2 - y_2\|_2^2. \quad (16)$$

We show that the first term on the right hand side of (16) is at most $O(\log^{3/2} d \sqrt{\log N \log(1/\delta)}) \text{opt}_{\varepsilon, \delta}(A, n)$ and the second term is $O(\log^2 d \log(1/\delta)) \text{opt}_{\varepsilon, \delta}(A, n)$.

The bound on $\mathbb{E}\|\tilde{y}_2 - y_2\|_2^2$ follows from Theorem 13. More precisely, $Y^T \tilde{y}_2$ is distributed identically to the output of $\mathcal{M}_g(Y^T A, x)$, and by Theorem 13,

$$\mathbb{E}\|\tilde{y}_2 - y_2\|_2^2 = \mathbb{E}\|Y^T \tilde{y}_2 - Y^T Ax\|_2^2 = O(\log^2 d \sqrt{\log N \log(1/\delta)}) \text{opt}_{\varepsilon, \delta}(A, \frac{d_{t+1}}{\varepsilon}).$$

Since, by the definition of t , $\frac{d_{t+1}}{\varepsilon} < n$, we have the desired bound.

The bound on $\mathbb{E}\|\hat{y}_1 - y_1\|_2^2$ follows from Lemma 1 and Lemma 9. We will use the notations for w_i , and r_i defined in Algorithm 2. Let $L = nXX^T K$; by Lemma 1,

$$\mathbb{E}\|\hat{y}_1 - y_1\|_2^2 \leq 4\mathbb{E}\|\hat{y}_1 - y_1\|_{L^\circ}^2 = 4\mathbb{E} \max_{j=1}^N |\langle na_j, XX^T(\tilde{y} - y) \rangle|.$$

The last equality follows from the definition of the dual norm $\|\cdot\|_{L^\circ}$ and from the fact that L is a polytope with vertices $\{a_j\}_{j=1}^N$, so any linear functional on L is maximized at one of the vertices. From the fact that for all $i \neq j$ we have $U_i^T U_j = 0$, from the triangle inequality, and from Lemma 9, we derive

$$\begin{aligned} \mathbb{E} \max_{j=1}^N |\langle na_j, XX^T(\tilde{y} - y) \rangle| &= \mathbb{E} \max_{j=1}^N n |\sqrt{k} \sum_{i=1}^t \langle a_j, r_i U_i w_i \rangle| \\ &\leq \sqrt{k} \sum_{i=1}^t \mathbb{E} \max_{j=1}^N n |\langle a_j, r_i U_i w_i \rangle| \\ &\leq O(\sqrt{kt} \sqrt{\log N \log(1/\delta)}) \text{opt}_{\varepsilon, \delta}(A, n) \\ &= O(\log^{3/2} d \sqrt{\log N \log(1/\delta)}) \text{opt}_{\varepsilon, \delta}(A, n). \end{aligned}$$

This completes the proof. \blacksquare

3.4 Computational Complexity

In this subsection we consider the computational complexity of our algorithms. We pay special attention to approximating the minimum enclosing ellipsoid of a polytope and computing least squares estimators. For both problems we need to go into the properties of known approximation algorithms in order to verify that the approximations are sufficient to guarantee that our algorithms can be implemented in polynomial time without hurting their near-optimality.

3.4.1 Minimum Enclosing Ellipsoid

Computationally the most expensive step of the base decomposition algorithm (Algorithm 1) is computing the minimum enclosing ellipsoid E of K . Computing the exact MEE can be costly: the fastest known algorithms have complexity on the order of $d^{O(d)} N$ [AS93]. However, for our purposes it is enough to compute an approximation of E in Banach-Mazur distance, i.e. some ellipsoid E' satisfying $\frac{1}{C}E' \subseteq E \subseteq E'$ for an absolute constant

C. Known approximation algorithms for MME guarantee that their output is an enclosing ellipsoid with volume approximately equal to that of the MEE [Kha96, TY07]. It is not immediately clear whether such an approximation is also a Banach-Mazur approximation. However, we can use the fact that the algorithms in [Kha96, KY05, TY07] output an ellipsoid E' satisfying approximate complimentary slackness conditions and show that $\Pi E'$ approximates ΠE in Banach-Mazur sense for some projection Π onto a subspace of dimension $\Omega(d)$. This suffices for a slightly weaker version of Lemma 5.

We begin with a definition. Let's define a vector $p \in [0, 1]^N$ to be C -optimal for $A = (a_i)_{i=1}^N$ if the following conditions are satisfied:

- $\sum_{i=1}^N p_i = 1$;
- for all $i \in [N]$, $a_i^T (APA^T)^{-1} a_i \leq C \cdot d$ where $P = \text{diag}(p)$ (we use this notation throughout this section).

The C -optimality conditions are a relaxation of the Karush-Kuhn-Tucker conditions of a formulation of the MEE problem as convex program. The approximation algorithm for MEE due to Khachiyan [Kha96], and later follow up work [KY05, TY07] compute a C -optimal p and output an approximate MEE ellipsoid $E(p) = \{x : x^T (APA^T)^{-1} x \leq Cd\}$. The running time of the algorithm in [TY07] is $\tilde{O}(d^2 N)$, where the \tilde{O} notation suppresses dependence on C as well as polylogarithmic terms.

C -optimality implies the following property of the the ellipsoid $E(p)$ which is key to our analysis.

Lemma 10 *Let $E^* = F^* B_2^d$ be the minimum enclosing ellipsoid of $K = \text{sym}\{a_i\}_{i=1}^N$, and let p be C -optimal for $A = (a_i)_{i=1}^N$. Let also $E(p) = \{x : x^T (APA^T)^{-1} x \leq Cd\} = F(p) B_2^d$. Then,*

$$\sigma_{d/2}^2(F(p)) \leq 4C \sigma_{d/4}^2(F^*).$$

Proof: Let $G = (F^*)^{-1}$. Since $GE^* = B_2^d$, we have that the MEE of GK is the unit ball, and, therefore, $\|Ga_i\|_2^2 \leq 1$ for all $i \in [N]$. Since $F(p)F(p)^T = Cd \cdot APA^T$, we have

$$\begin{aligned} \frac{1}{d} \sum_{i=1}^N \sigma_i^2(GF(p)) &= \frac{1}{d} \text{tr}(GF(p)F(p)^T G^T) \\ &= C \text{tr}(GAPA^T G^T) \\ &= C \sum_{i=1}^N p_i \|Ga_i\|_2^2 \leq C. \end{aligned}$$

By Markov's inequality, $\sigma_{d/4}^2(GF(p)) \leq 4C$. Let Π_1 be the projection operator onto the subspace spanned by the left singular vectors of $GF(p)$ corresponding to $\sigma_{d/4}(GF(p)), \dots, \sigma_d(GF(p))$. We have $\Pi_1 GE(p) \subseteq 2C^{1/2} \Pi_1 B_2^d$. Multiplying on both sides by F^* , we get

$$F^* \Pi_1 GE(p) \subseteq 2C^{1/2} F^* \Pi_1 B_2^d.$$

Let Π_2 be the matrix $\Pi_2 = G^{-1} \Pi_1 G = F^* \Pi_1 G$. Since Π_2 is similar to Π_1 , it is also a projection matrix onto a $3d/4$ dimensional subspace. We have that $F^* \Pi_1 = \Pi_2 F^*$, and therefore

$$\Pi_2 E(p) \subseteq 2C^{1/2} \Pi_2 E^*.$$

Define $H = 4CF^*(F^*)^T - F(p)F(p)^T$. The inclusion above is equivalent to the positive semidefiniteness of the matrix $\Pi_2 H \Pi_2^T$. As Π_2 is a projection onto a $3d/4$ dimensional subspace, by the standard minimax characterization of eigenvalues we have $\lambda_{3d/4}(H) \geq 0$. We recall the (dual) Weyl inequalities for symmetric $d \times d$ matrices X and Y :

$$\lambda_i(X) + \lambda_j(Y) \leq \lambda_{i+j-d}(X + Y).$$

The inequalities are standard and follow from the minimax characterization of eigenvalues and dimension counting arguments — see, e.g. Chapter 1 in [Tao12]. Substituting $X = H$ and $Y = F(p)F(p)^T$, $i = 3d/4$ and $j = d/2$, we have the inequality

$$\sigma_{d/2}^2(F(p)) = \lambda_{d/2}(F(p)F(p)^T) \leq \lambda_{d/4}(4CF^*(F^*)^T) = 4C \sigma_{d/4}^2(F^*),$$

and the proof is complete. ■

Finally we give an analogue of Lemma 5 for the variant of the base decomposition algorithm that uses an approximate MEE. The proof follows from Lemma 10 and the arguments used to prove Lemma 5. We omit a full proof here.

Lemma 11 *Consider a variant of Algorithm 1 that, at each step, uses $E(p) = \{x : x^T(APA^T)^{-1}x \leq Cd\} = F(p)B_2^d$, where p is $O(1)$ -optimal for A , rather than the minimum enclosing ellipsoid $E = FB_2^d$. Let d_i be the dimension of the span of U_i . For any i there exists a subspace W_i of dimension $\Omega(d_i)$ and a set $S_i \subseteq [N]$ of size $|S_i| = \Omega(d_i)$, such that $\sigma_{\min}^2(W_i W_i^T A|_{S_i}) = \Omega(1) \max_{j=1}^N \|U_i^T a_j\|_2^2$.*

One can verify that in all our proofs we can substitute Lemma 11 for Lemma 5 without changing the asymptotics of our lower and upper bounds. Therefore, in all our algorithms we can use the variant of Algorithm 1 from Lemma 11 without compromising near-optimality. This variant of Algorithm 1 runs in time $d^{O(1)}N$.

Notice that the base decomposition can be reused for different databases, as long as the query matrix A stays unchanged; once the decomposition is computed the rest of the algorithm is very efficient: it involves some standard algebraic computations and sampling from an $O(d)$ -dimensional gaussian distribution. Furthermore, any ellipsoid E' containing K suffices for privacy, and one may use heuristic approximations to the MEE problem.

3.4.2 Least Squares Estimator

Except for base decomposition, the other potentially computationally expensive step in Algorithm 3 is the computation of a least squares estimator \hat{y}_1 . This is a quadratic minimization problem, and can be approximated by the simple Frank-Wolfe gradient descent algorithm [FW56]. In particular, for a point \hat{y}' such that $\|\hat{y}' - y\|_2^2 \leq \min_{\hat{y} \in L} \|\hat{y}' - y\|_2^2 + \alpha$, Lemma 1 holds to within an additive approximation factor α , i.e. $\|\hat{y}' - y\|_2^2 \leq 4\|w\|_{L^\circ} + \alpha$. We call such a point \hat{y}' an α additive approximation to the least squares estimator problem. By the analysis of Clarkson [Cla10], T iterations of the Frank-Wolfe algorithm give an additive approximation where $\alpha \leq 4C(L)/(T+3)$, for $C(L) \leq \sup_{u,v \in L} |\langle u, u-v \rangle|$. In our case $L = nXX^TK$. In order to have near optimality for Algorithm 3, an additive approximation $\alpha \leq \sum_{i=1}^t nr_i^2$ suffices. Using the triangle inequality and Cauchy-Schwarz, we can bound $C(L)$ for $L = nXX^TK$ as

$$C(L) \leq \sum_{i=1}^t \sup_{u,v \in nU_i U_i^T K} \langle u^T, u-v \rangle \leq \sum_{i=1}^t 2n^2 r_i^2.$$

Therefore, $T = O(n)$ iterations of the Frank-Wolfe algorithm suffice. Since each iteration of the algorithm involves N dot product computations and solving a homogeneous linear system in at most d variables and at most d equations, it follows that an approximate version of Algorithm 3 with unchanged asymptotic optimality guarantees can be implemented in time $d^{O(1)}Nn$.

We note that the approximation algorithm of Khachiyan for the MEE problem, as well as its modification in [TY07], can also be interpreted as instances of the Frank-Wolfe algorithm (see [TY07] for details).

4 Results for Pure Privacy

Our geometric approach to approximate privacy allows us to better understand the optimal error required for approximate privacy vs. that required for pure privacy. Our first result bounds the gap between the optimal error bounds for the two notions of privacy in the dense case. Then we extend these ideas and give a $(\epsilon, 0)$ -differentially private algorithm which nearly matches the guarantees of Algorithm 3 for sparse databases.

4.1 The Cost of Pure Privacy

In this subsection we investigate the worst-case gap between $\text{opt}_{\epsilon, \delta}(A)$ (for small enough $\delta > 0$) and $\text{opt}_{\epsilon, 0}(A)$ over all query matrices A . At the core of our analysis is a natural geometric fact: for any symmetric polytope K with N vertices in d -dimensional space we can find a subset of d vertices of K whose symmetric convex hull has volume radius at most a factor $O(\sqrt{\log(N/d)})$ smaller than the volume radius of K . Our proof of this fact goes through analyzing the contact points of K with its minimum enclosing volume ellipsoid, and a bound on the volume of polytopes with few vertices.

Theorem 15 *Let $K = \text{sym}\{a_1, \dots, a_N\} \subseteq \mathbb{R}^d$ and let E be an ellipsoid of minimal volume containing K . There exists a set $S \subseteq [N]$ of size d such that the matrix $A|_S = (a_i)_{i \in S}$ satisfies $\det(A|_S)^{1/d} = \Omega(\text{vrad}(E))$.*

For the proof of 15 we will use John's theorem (Theorem 6) and the following elementary algebraic result.

Lemma 12 *Let u_1, \dots, u_m be d -dimensional unit vectors and let c_1, \dots, c_m be positive reals such that*

$$\sum c_i u_i u_i^T = I. \quad (17)$$

Then there exists a set $S \subseteq [m]$ of size d such that the matrix $U = (u_i)_{i \in S}$ satisfies $|\det(U)|^{1/d} = \Omega(1)$.

Proof: Notice that $\text{tr}(u_i u_i^T) = \|u_i\|_2^2 = 1$ for all i . By taking traces of both sides of (17), we have $\sum c_i = d$.

Let $U = (u_i)_{i=1}^m$ and let C be the $m \times m$ diagonal matrix with $(c_i)_{i=1}^m$ on the diagonal. Then we can write $I = \sum c_i u_i u_i^T = UCU^T$. By the Binet-Cauchy formula for the determinant,

$$\begin{aligned} 1 = \det(UCU) &= \sum_{S \subseteq [m]: |S|=d} \prod_{i \in S} c_i \det(U|_S)^2 \\ &\leq \max_{S \subseteq [m]: |S|=d} \det(U|_S)^2 \sum_{S \subseteq [m]: |S|=d} \prod_{i \in S} c_i \\ &\leq \max_{S \subseteq [m]: |S|=d} \det(U|_S)^2 \frac{1}{d!} \left(\sum_{i=1}^m c_i \right)^d \\ &\leq \max_{S \subseteq [m]: |S|=d} \det(U|_S)^2 \frac{d^d}{d!} \end{aligned} \quad (18)$$

The inequality (18) follows since each term $\prod_{i \in S} c_i$ appears $d!$ times in the expansion of $(\sum c_i)^d$ and all other terms in the expansion are positive. Using the inequality $d! \geq (d/e)^d$, we have that $\max_{S \subseteq [m]: |S|=d} \det(U|_S)^{2/d} \geq 1/e$, and this completes the proof. ■

Proof of Theorem 15: We can write the minimum enclosing ellipsoid E as $\text{vrad}(E)FB_2$ where F is a linear map with determinant 1. Since F^{-1} does not change volumes, B_2 is a minimal volume ellipsoid of $\text{vrad}(E)^{-1}F^{-1}K$. Also, for any $A|_S = (a_i)_{i \in S}$, where $S \subseteq [N]$, we have

$$\det(A|_S) = \text{vrad}(E)^d \det(\text{vrad}(E)^{-1}F^{-1}A|_S).$$

Therefore, it is sufficient to show that for $L = \text{sym}\{u_1, \dots, u_N\}$ such that B_2 is the minimal volume ellipsoid of L , there exists a set $S \subseteq [N]$ such that the matrix $U|_S = (u_i)_{i \in S}$ satisfies $\det(U|_S)^{1/d} = \Omega(1)$. Since, by convexity, the contact points $L \cap B_2$ of L are a subset of u_1, \dots, u_N , the statement follows from Theorem 6 and Lemma 12. ■

Combined with the following theorem of Bárány and Füredi [BF88], and also Gluskin [Glu07] (with sharper bounds), Theorem 15 implies the corollary that for any d -dimensional symmetric polytope one can find a set of d vertices whose symmetric convex hull captures a significant fraction of the volume of the polytope.

Theorem 16 ([BF88, Glu07]) *Let $K = \text{sym}\{a_1, \dots, a_N\}$ and let E be an ellipsoid containing K . Then $\text{vrad}(K) \leq O\left(\sqrt{\frac{\log(N/d)}{d}}\right) \text{vrad}(E)$.*

Corollary 1 *For any $K = \text{sym}\{a_1, \dots, a_N\}$ there exists a set $S \subseteq [N]$ such that*

$$\det((a_i)_{i \in S})^{1/d} = \Omega\left(\sqrt{\frac{d}{\log(N/d)}}\right) \text{vrad}(K).$$

Finally, we describe the application to differential privacy. By Corollary 1, $\text{volLB}(A, \varepsilon) = O(\frac{1}{\varepsilon^2} \log(N/d)) \det \text{LB}(A, d)$. Also, by (11), $\det \text{LB}(A, d) = O(\log^2 d) \text{opt}_{c_1, c_2}(A, d)$. Finally, Lemma 3 implies that $\text{opt}_{c_1, c_2}(A, d) \leq \varepsilon^2 \text{opt}_{\varepsilon, \delta}(A, d/\varepsilon)$ for δ small enough with respect to ε . Putting all this together and using Theorem 12, we have the following theorem (Theorem 2).

Theorem 17 *For small enough ε and all δ small enough with respect to ε , for any $d \times N$ real matrix A we have*

$$\text{opt}_{\varepsilon, 0}(A) = O(\log^3 d) \text{volLB}(A, \varepsilon) = O(\log^5 d \log(N/d)) \text{opt}_{\varepsilon, \delta}(A, \frac{d}{\varepsilon}).$$

4.2 Sparse Case under Pure Privacy

We further extend our results from Section 3.3 and show an efficient $(\varepsilon, 0)$ -differentially private algorithm which, on input any query matrix A and any database size bound n , nearly matches $\text{opt}_{\varepsilon,0}(A, n)$. This proves our main Theorem 4. In fact, our result is stronger: we show an $(\varepsilon, 0)$ -differentially private mechanism whose error nearly matches $\text{opt}_{\varepsilon,\delta}(A, n)$ for all δ small enough with respect to ε . Thus, the result of this subsection can be seen as a generalization of Theorem 17 to the sparse databases regime.

Our algorithm for sparse databases under pure privacy closely follows Algorithm 3: we add noise from a distribution that is tailored to A but oblivious to the database x ; then we use least squares estimation to reduce error on sparse databases. However, Gaussian noise does not preserve $(\varepsilon, 0)$ -differential privacy, and we need to use a different noise distribution. Intuitively, one expects that adding noise sampled from a near-optimal distribution [HT10, BDKT12] and then computing a least squares estimator would be nearly optimal, analogously to Algorithm 3. We are not currently able to analyze the error of this algorithm, but instead we analyze a variant of Algorithm 3 where the Gaussian distribution is simply substituted with the generalized K -norm distribution from [HT10]. Intuitively, we are able to show that the generalized K -norm distribution “approximates a Gaussian” well enough for our analysis to go through. A main tool in our analysis is a classical concentration of measure inequality from convex geometry.

We begin with a slight generalization of the main upper bound result of Hardt and Talwar [HT10]. This generalization follows directly from the methods used in [HT10] with only minor modifications in the proofs. We omit a full derivation here. Also, while the methods of Hardt and Talwar will lead to a proof conditional on the truth of the Hyperplane conjecture from convex geometry, using the ideas of Bhaskara et al. [BDKT12] the result can be made unconditional.

Theorem 18 ([HT10, BDKT12]) *Let $A = (a_i)_{i=1}^N$ be an $d \times N$ real matrix and let $K = \text{sym}\{a_i\}_{i=1}^N$. There exists an efficiently computable and efficiently sampleable distribution $\mathcal{W}(A, \varepsilon)$ such that the following claims hold:*

1. *the algorithm \mathcal{M}_K which on input x outputs $Ax + w$ for $w \sim \mathcal{W}(A, \varepsilon)$ satisfies $(\varepsilon, 0)$ -differential privacy;*
2. *$\mathcal{W}(A, \varepsilon)$ is identical to the distribution of the random variable $m \sum_{\ell=1}^m w_\ell$ where $m = O(\log d)$, and w_ℓ is a sample from a log concave distribution with support lying in a subspace \mathcal{V}_ℓ of \mathbb{R}^d of dimension d_ℓ ;*
3. *\mathcal{V}_ℓ and $\mathcal{V}_{\ell'}$ for $\ell' \neq \ell$ are orthogonal, and the union of $\{\mathcal{V}_\ell\}_{\ell=1}^m$ spans \mathbb{R}^d ;*
4. *let $M_\ell = M_\ell(A, \varepsilon) = \mathbb{E} m^2 w_\ell w_\ell^T$ be the correlation matrix of $m w_\ell$ and let Π_ℓ be the projection matrix onto $\text{span}\{\mathcal{V}_j\}_{j=\ell}^m$; then*

$$\lambda_{\max}(M_\ell) \leq O(\log^2 d) \frac{d_\ell}{\varepsilon^2} \text{vrad}(\Pi_\ell K)^2,$$

where $\lambda_{\max}(M_\ell)$ is the largest eigenvalue of M_ℓ .

Using the distribution $\mathcal{W}(A, \varepsilon)$, we define our near optimal sparse-case algorithm satisfying pure differential privacy as Algorithm 4.

Algorithm 4 LEAST SQUARES MECHANISM: PURE PRIVACY

Input (*Public*): query matrix $A = (a_i)_{i=1}^N \in \mathbb{R}^{d \times N}$ ($\text{rank} A = d$); database size bound n

Input (*Private*): database $x \in \mathbb{R}^N$

Let U_1, \dots, U_k be base decomposition computed by Algorithm 1 on input A , where U_i is an orthonormal basis for a space of dimension d_i ;

Let t be the largest integer such that $d_t \geq \varepsilon n$;

for all $i \leq t$ **do**

 Compute $\tilde{y}_i = U_i(U_i^T Ax + w_i)$ where $w_i \sim \mathcal{W}(U_i^T A, \frac{\varepsilon}{t+1})$.

end for

Let $\tilde{y}' = \sum_{i=1}^t \tilde{y}_i$

Let $X = \sum_{i=1}^t U_i$ and $Y = \sum_{i=t+1}^k U_i$;

Compute $\tilde{y}'' = Y(Y^T Ax + w'')$ where $w'' \sim \mathcal{W}(Y^T A, \frac{\varepsilon}{t+1})$;

Compute $\hat{y}' = \arg \min \{\|\tilde{y}' - \hat{y}'\|_2^2 : \hat{y}' \in nXX^T K\}$, where $K = AB_1$.

Output $\hat{y}' + \tilde{y}''$.

Theorem 19 Let $\mathcal{M}_p(A, x, n)$ be the output of Algorithm 4 on input a $d \times N$ query matrix A , database size bound n and private input x . $\mathcal{M}_p(A, x, n)$ is $(\varepsilon, 0)$ -differentially private and for all small enough ε and all δ small enough with respect to ε satisfies

$$\begin{aligned}\text{err}_{\mathcal{M}_p}(A) &= O(\log^4 d \log^{3/2} N) \text{opt}_{\varepsilon, \delta}(A, n) + O(\log^5 d \log N) \text{opt}_{\varepsilon, 0}(A, n); \\ \text{err}_{\mathcal{M}_p}(A) &= O(\log^4 d \log^{3/2} N + \log^7 d \log N) \text{opt}_{\varepsilon, \delta}(A, n).\end{aligned}$$

Once again privacy follows by a straightforward argument from the privacy of the underlying noise-adding mechanism, in this case the generalized K -norm mechanism.

Lemma 13 $\mathcal{M}_p(A, x, n)$ satisfies $(\varepsilon, 0)$ -differential privacy.

Proof: $\mathcal{M}_p(A, x, n)$ is a deterministic function of $\tilde{y}_1, \dots, \tilde{y}_t$ and \tilde{y}'' . Each of these quantities is the output of an algorithm satisfying $(\frac{\varepsilon}{t+1}, 0)$ -differential privacy (by Theorem 18, claim 1). Therefore, by Lemma 2, $\mathcal{M}_p(A, x, n)$ satisfies $(\varepsilon, 0)$ -differential privacy. \blacksquare

Next we prove the main technical lemma we need in order to show near optimality. The analysis is very similar to that of Lemma 9. The main technical challenge is to show that the distribution \mathcal{W} has all the properties we needed from the Gaussian distribution: covariance with bounded operator norm and exponential concentration. We use ideas from Section 4.1 and the following variant of a classical concentration of measure inequality, due to Borell [Bor75] (proved in the appendix).

Theorem 20 Let μ be a log-concave distribution over \mathbb{R}^d . Assume that A is a symmetric convex subset of \mathbb{R}^d such that $\mu(A) = \theta \geq \frac{2}{3}$. Then, for every $t > 1$ we have

$$\mu[(tA)^c] \leq 2^{-(t+1)/2}$$

We are now ready to prove the counterpart of Lemma 9 for \mathcal{M}_p .

Lemma 14 Let U_i , t , and w_i be as defined in Algorithm 4. For all small enough ε and all δ small enough with respect to ε , $\mathbb{E} \max_{j=1}^N n |\langle a_j, \sum_{i=1}^t U_i w_i \rangle| \leq O(\log^4 d \log^{3/2} N) \text{opt}_{\varepsilon, \delta}(A, n)$.

Proof: As in Algorithm 3, we define $r_i = \max_{j=1}^N \|U_i^T a_j\|_2$. Equivalently r_i is the radius of the smallest d_i -dimensional ball which contains $U_i^T K$. In the proof of Lemma 9 we argued that $\text{opt}_{\varepsilon, \delta}(A, n) = \Omega(\frac{n}{\varepsilon} r_i^2)$ for all small enough ε and all δ small enough with respect to ε .

By Theorem 18, we can write w_i as $m_i \sum_{\ell=1}^{m_i} w_{i\ell}$ where $w_{i\ell}$ is a sample from a log concave distribution over a subspace $\mathcal{V}_{i\ell}$ and $m_i = O(\log d_i)$. Furthermore, all $w_{i\ell}$ for a fixed i are mutually orthogonal. Finally, letting $\Pi_{i\ell}$ be the projection matrix onto $\text{span}\{\mathcal{V}_{i\ell}\}_{\ell=1}^{m_i}$ and $M_{i\ell}$ be the covariance matrix of $m_i w_{i\ell}$, we have

$$\lambda_{\max}(M_{i\ell}) \leq O(\log^2 d) \frac{d_{i\ell} t^2}{\varepsilon^2} \text{vrad}(\Pi_{i\ell} U_i^T K)^2. \quad (19)$$

Therefore, for any a_j , we can derive the following bound:

$$\begin{aligned}\mathbb{E} n^2 |\langle a_j, U_i w_i \rangle|^2 &= n^2 \mathbb{E} |\langle U_i^T a_j, w_i \rangle|^2 \\ &\leq n^2 m_i \sum_{\ell=1}^{m_i} \mathbb{E} |\langle U_i^T a_j, m_i w_{i\ell} \rangle|^2 \\ &= n^2 m_i \sum_{\ell=1}^{m_i} (U_i^T a_j)^T M_{i\ell} (U_i^T a_j) \\ &\leq n^2 m_i \sum_{\ell=1}^{m_i} \|U_i^T a_j\|_2^2 \lambda_{\max}(M_{i\ell}) \\ &\leq O(\log^5 d) n^2 \sum_{\ell=1}^{m_i} \frac{1}{\varepsilon^2} r_i^2 d_{i\ell} \text{vrad}(\Pi_{i\ell} U_i^T K)^2.\end{aligned}$$

The first bound above follows from the Cauchy-Schwarz inequality and the last bound follows from (19). To bound $d_{i\ell} \text{vrad}(\Pi_{i\ell} U_i^T K)^2$, recall that $U_i^T K$ is contained in a ball of radius r_i , and therefore so is $\Pi_{i\ell} U_i^T K$ for all

ℓ . Therefore, by Theorem 16, $\text{vrad}(\Pi_{i\ell} U_i^T K)^2 = O((\log(N/d_{i\ell}))/d_{i\ell}) r_i^2$. Substituting into the bound above and recalling that $\text{opt}_{\varepsilon,\delta}(A, n) = \Omega(\frac{n}{\varepsilon} r_i^2)$, we get

$$\mathbb{E} n^2 |\langle a_j, U_i w_i \rangle|^2 = O(\log^6 d \log(N/d)) \text{opt}_{\varepsilon,\delta}(A, n)^2.$$

Thus applying Cauchy Schwarz once again, we conclude

$$\mathbb{E} n^2 |\langle a_j, \sum_{i=1}^t U_i w_i \rangle|^2 \leq t \cdot \sum_{i=1}^t \mathbb{E} n^2 |\langle a_j, U_i w_i \rangle|^2 = O(\log^8 d \log(N/d) \text{opt}_{\varepsilon,\delta}(A, n)).$$

For any j , the set $\{w : |\langle U_i^T a_j, \sum_i w_i \rangle| \leq T\}$ is symmetric and convex for any bound T . Then, by Chebyshev's inequality, and Theorem 20 there exists a constant C such that for any i, j and $\alpha > 2$

$$\Pr[n |\langle U_i^T a_j, w_i \rangle| > C \alpha \log N \log^4 d \sqrt{\log(N/d)} \frac{n}{\varepsilon} r_i^2] < N^{-\alpha}.$$

Using a union bound and taking expectations completes the proof. \blacksquare

Proof of Theorem 19: The privacy guarantee follows from Lemmas 13. By Lemma 14 analogously to the proof of Theorem 14, we can conclude that

$$\mathbb{E}[\|\hat{y}' - X X^T A x\|_2^2] \leq 4 \mathbb{E} \max_{j=1}^N n |\langle a_j, \sum_{i=1}^t U_i w_i \rangle| \leq O(\log^4 d \log^{3/2} N) \text{opt}_{\varepsilon,\delta}(A, n).$$

Moreover, by Theorem 12,

$$\begin{aligned} \mathbb{E}[\|\tilde{y}'' - Y Y^T A x\|_2^2] &\leq O(\log^3 d) \text{volLB}(Y Y^T A, \frac{\varepsilon}{t+1}) \\ &= O(t^2 \log^3 d) \text{opt}_{\varepsilon,0}(Y Y^T A, \frac{d_t}{\varepsilon}) \\ &= O(\log^5 d) \text{opt}_{\varepsilon,0}(A, n). \end{aligned}$$

The last bound follows since $Y Y^T$ is a projection matrix, $d_t \leq n\varepsilon$ and $t = O(\log d)$. Also, by Theorem 17, $\text{volLB}(Y Y^T A, \frac{\varepsilon}{t+1}) = O(t^2 \log^2 d \log(N/d_t)) \text{opt}_{\varepsilon,\delta}(Y Y^T A, \frac{d_t}{\varepsilon})$, and therefore we have $\mathbb{E}[\|\tilde{y}'' - Y Y^T A x\|_2^2] = O(\log^7 d \log(N/d)) \text{opt}_{\varepsilon,\delta}(A, n)$.

Pythagoras theorem then implies the result. \blacksquare

5 Universal bounds

For d linear sensitivity 1 queries, there are known universal bounds on $\text{err}_{\varepsilon,\delta}(A, n)$ and $\text{err}_{\varepsilon,0}(A, n)$. We note that the sensitivity 1 bound implies that the ℓ_2 norm of each column is at most \sqrt{d} . This in turn allows us to prove an upper bound on the spectral lower bound, so that the relative guarantee provided by Theorems 13 and 14 can be translated to an absolute one. The resulting bounds can be improved by polylogarithmic factors, by using natural simplifications of the relative-error mechanisms and analyzing them directly. We next present these simplifications. The average per query error bounds resulting from our mechanisms match the best known bounds for (ε, δ) -differential privacy, and improve on the best known bounds for pure differential privacy.

For (ε, δ) -differential privacy, the best known universal upper bound when $A \in [0, 1]^{d \times N}$ for the total ℓ_2^2 error is $O(nd \log d \sqrt{\log N} \log(1/\delta)/\varepsilon)$, given by [GRU12]. We note that when $A \in [0, 1]^{d \times N}$, one can use $B(0, \sqrt{d})$ as an enclosing ellipsoid for K . The following simple mechanism is easily seen to be (ε, δ) -DP.

Theorem 21 *The mechanism \mathcal{M}_s of Algorithm 5 is (ε, δ) -differentially private and satisfies*

$$\text{err}_{\mathcal{M}_s}(A) \leq O(nd \log(1/\delta) \sqrt{\log N}/\varepsilon)$$

Proof: The privacy of the mechanism is immediate from Lemma 4. To analyze the error, we use Lemma 1 and the fact that $L = nK$ to bound

$$\text{err}_{\mathcal{M}_s}(A) = \mathbb{E}[\|\hat{y} - A x\|_2^2] \leq 4n \|w\|_{K^\circ} = 4n \mathbb{E} \max_{j=1}^N |\langle a_j, w \rangle|,$$

Algorithm 5 SIMPLE NOISE + LEAST SQUARES MECHANISM

Input Public Input: query matrix $A = (a_i)_{i=1}^N \in [0, 1]^{d \times N}$

Input Private Input: database $x \in \mathbb{R}^N$

Let $c(\varepsilon, \delta) = \frac{1 + \sqrt{2 \ln(1/\delta)}}{\varepsilon}$.

Let $r = \max_{j=1}^N \|a_j\|_2$;

Sample $w \sim N(0, c(\varepsilon, \delta))^d$;

Let $\tilde{y} = Ax + rw$.

Let $\hat{y} = \arg \min\{\|\tilde{y} - \hat{y}\|_2^2 : \hat{y} \in nK\}$, where $K = AB_1$.

Output \hat{y}

where $\{a_j\}_{j=1}^N$ are columns of A . We have used the fact that the $\|w\|_{K^\circ} = \max_{a \in K} \langle a, w \rangle$ is attained at one of the vertices of K . Since each $\langle a_j, w \rangle$ is a Gaussian with variance $r^4 c(\varepsilon, \delta)^2$, $|\langle a_j, w \rangle|$ exceeds $r^2 c(\varepsilon, \delta) \sqrt{t \log N}$ with probability at most $\frac{1}{N^t}$. Taking a union bound, we conclude that this expectation of the maximum is $O(r^2 c(\varepsilon, \delta) \sqrt{\log N})$. Recall that $r = \max_{j=1}^N \|a_j\|_2 \leq \sqrt{d}$. It follows that

$$\text{err}_{\mathcal{M}_s}(A) \leq O(ndc(\varepsilon, \delta) \sqrt{\log N})$$

■

Comparing with [GRU12], our bound is better by an $O(\log d)$ factor. However, the previous bound is stronger in that it guarantees expected squared error $O_{\varepsilon, \delta}(n \sqrt{\log N} \log d)$ for every query, while we can only bound the total ℓ_2^2 error.

For getting pure ε -DP, we simply substitute the generalized K -norm distribution guaranteed by Theorem 18 instead of the Gaussian noise.

Algorithm 6 K -NORM NOISE + LEAST SQUARES MECHANISM

Input Public Input: query matrix $A = (a_i)_{i=1}^N \in [0, 1]^{d \times N}$

Input Private Input: database $x \in \mathbb{R}^N$

Let $\tilde{y} = Ax + w$ where $w \sim \mathcal{W}(A, \varepsilon)$.

Let $\hat{y} = \arg \min\{\|\tilde{y} - \hat{y}\|_2^2 : \hat{y} \in nK\}$, where $K = AB_1$.

Output \hat{y}

We first observe an upper bound on the volume radius of projections of K .

Lemma 15 *Let $A \in [0, 1]^{d \times N}$ and let $K = \text{sym}\{a_1, \dots, a_N\}$. Let $\Pi^{(k)}$ be a rank k orthogonal projection that maps \mathbb{R}^d to \mathbb{R}^k . Then*

$$\text{vrad}(\Pi^{(k)} K) \leq O\left(\sqrt{\frac{\log(N/k)}{k}}\right) \sqrt{d}$$

Proof: Since each column of A has ℓ_2 norm at most \sqrt{d} , $\Pi^{(k)} K$ is contained in a ball of radius \sqrt{d} . Theorem 16 then immediately implies the claimed bound. ■

Now we show that Algorithm 6 achieves the bound claimed in Theorem 5.

Theorem 22 *The mechanism \mathcal{M}_{sp} of Algorithm 6 is $(\varepsilon, 0)$ -differentially private and satisfies*

$$\text{err}_{\mathcal{M}_{sp}}(A) \leq O(nd\varepsilon^{-1} \log^2 d \log^{\frac{3}{2}} N)$$

Proof: The privacy property follows from Theorem 18 and the fact that post-processing preserves $(\varepsilon, 0)$ -differential privacy. To prove the error bound, we need to upper bound $\mathbb{E}_w[\|w\|_{K^\circ}]$ for a polytope $K \subseteq \mathbb{R}^d$ with N vertices, where w is drawn from $\mathcal{W}(A, \varepsilon)$. It therefore suffices to bound $\mathbb{E}_w[\max_{i=1}^N |\langle a_i, w \rangle|]$.

By Theorem 18, we can write w as $\sum_{\ell=1}^m m w_\ell$ where w_ℓ is drawn from a log concave distribution and $m = O(\log d)$.

For a fixed ℓ ,

$$\begin{aligned}\mathbb{E}[|\langle a_i, mw_\ell \rangle|^2] &= a_i^T M_\ell a_i \\ &\leq \|a_i\|^2 \cdot \lambda_{\max}(M_\ell) \\ &\leq O(d \log^2 d) \frac{1}{\varepsilon^2} d_\ell \text{vrad}(\Pi_\ell K)^2.\end{aligned}$$

By lemma 15, this is at most $O(d \log^2 d \log(N/d_\ell) d / \varepsilon^2) \leq O(\frac{d^2}{\varepsilon^2} \log^2 d \log N)$. Then by Cauchy-Schwarz,

$$\begin{aligned}\mathbb{E}[|\langle a_i, \sum_{\ell=1}^m mw_\ell \rangle|^2] &\leq m \cdot \sum_{\ell=1}^m \mathbb{E}[|\langle a_i, mw_\ell \rangle|^2] \\ &\leq O\left(\frac{d^2}{\varepsilon^2} \log^4 d \log N\right)\end{aligned}$$

By Theorem 20, $\Pr[|\langle a_i, \sum_{\ell} w_\ell \rangle| \geq t \log N \cdot O(\frac{d}{\varepsilon} \log^2 d \sqrt{\log N})] \leq N^{-\Omega(t)}$. It follows that $\mathbb{E}_w[\|w\|_{K^\circ}]$ is bounded by $O(\frac{d}{\varepsilon} \log^2 d \log^{\frac{3}{2}} N)$.

It then follows that $\mathbb{E}[\|\hat{y} - Ax\|_2^2]$ is $O(nd\varepsilon^{-1} \log^2 d \log^{\frac{3}{2}} N)$. ■

6 Extensions

In this section we describe a couple of extensions to our results. We show how to translate our optimality guarantees for total squared error in the dense case regime to worst case error over queries using the minimax theorem. We further show that our nearly optimal efficient mechanism in the dense case regime implies a polylogarithmic approximation to hereditary discrepancy.

6.1 Expected ℓ_∞ Error

For $i \in [d]$, let $a^{(i)}$ denote the i th row of A and let $\mathcal{M}(x)_i$ denote the i th coordinate of the answer $\mathcal{M}(x)$. Let $\text{err}_{\mathcal{M}}^{\ell_\infty}(A) = \max_{x \in \mathbb{R}^N} \mathbb{E}[\|Ax - \mathcal{M}(x)\|_\infty]$ denote the worst case (over databases) expected ℓ_∞ error of a mechanism for a query A . Here as usual, the expectation is taken over the internal coin tosses of \mathcal{M} . Thus we measure the expected worst case error $\mathbb{E}[\max_i |\langle a^{(i)}, x \rangle - \mathcal{M}(x)_i|]$. Let $\text{opt}^{\ell_\infty}(A) = \min_{\mathcal{M}: \mathcal{M} \text{ is } (\varepsilon, \delta)\text{-DP}} \text{err}_{\mathcal{M}}^{\ell_\infty}(A)$ denote the minimum ℓ_∞ error over all (ε, δ) -differentially private mechanisms. In the dense case, our results can be extended to the ℓ_∞ error at the cost of an additional $O(\log d)$ loss in the competitive ratio.

We derive such an extension in two steps. First, we give a mechanism for which the worst case expected squared error $\max_i \mathbb{E}[|\langle a^{(i)}, x \rangle - \mathcal{M}(x)_i|^2]$ is small. We then use this mechanism as a blackbox to derive one which has small expected ℓ_∞ error. For a mechanism \mathcal{M} , let $E_i^{\mathcal{M}, A} = \mathbb{E}[|\langle a^{(i)}, x \rangle - \mathcal{M}(x)_i|^2]$ denote the expected squared error of the mechanism in the i th coordinate.

Theorem 23 *Let $A \in \mathbb{R}^{d \times N}$, ε, δ be given. There is an (ε, δ) -DP mechanism \mathcal{M}_{we} , and a non-negative diagonal matrix P with $\text{tr}(P^2) = 1$ such that for all $i \in [d]$,*

$$E_i^{\mathcal{M}_{we}, A} \leq O(\log^2 d \log 1/\delta) \text{specLB}(PA) \leq O(\log^2 d \log 1/\delta) (\text{opt}^{\ell_\infty}(A))^2$$

Proof: Recall that for any A , the mechanism \mathcal{M}_g^A of Section 3.2 is (ε, δ) -DP and has total expected ℓ_2^2 error at most $\beta \triangleq O(\log^2 \log 1/\delta)$ times the lower bound $\text{specLB}(A)$. We will use this mechanism as a subroutine to derive \mathcal{M}_{we} .

Let (p_1, p_2, \dots, p_d) denote a probability distribution so that $p_i \geq 0$ and $\sum_i p_i = 1$. Let P denote a diagonal matrix with entries $\sqrt{p_i}$. It is easy to see that $\text{err}_{\mathcal{M}}^{\ell_2^2}(PA) = \sum_i p_i E_i^{\mathcal{M}, A}$. It follows that $\text{opt}_{\varepsilon, \delta}^{\ell_2^2}(PA) \leq (\text{opt}^{\ell_\infty}(A))^2$: indeed the mechanism \mathcal{M} achieving $\text{opt}^{\ell_\infty}(A)$ gives this error bound.

Thus the mechanism \mathcal{M}_g^{PA} satisfies

$$\text{err}_{\mathcal{M}_g^{PA}}^{\ell_2^2}(PA) \leq \beta \text{specLB}(PA) \leq \beta \text{opt}^{\ell_2^2}(PA).$$

It follows that $\sum_i p_i E_i^{\mathcal{M}_g^{PA}, A}$ is $\beta \cdot (\text{opt}^{\ell_\infty}(A))^2$. Consider a two-player zero sum game where the row-player selects a query $a^{(i)}$ from A , and the column player picks an (ε, δ) -DP mechanism \mathcal{M} and must pay the row player

$E_i^{\mathcal{M},A}$. The above discussion shows that for any randomized strategy (given by P) of the row player, the column player has a strategy that guarantees payoff at most $\beta \cdot (\text{opt}^{\ell_\infty}(A))^2$. By the minimax theorem, this also upper bounds the value of the game. Moreover, using standard constructive versions of the minimax theorems (see e.g. [AHK12]), one can come up with such a distribution \mathcal{D} over $O(d \log d)$ (ε, δ) -DP mechanisms, and a P such that for all $i \in [d]$,

$$\mathbb{E}_{\mathcal{M} \sim \mathcal{D}}[E_i^{\mathcal{M},A}] \leq 2\beta \text{specLB}(PA) \leq 2\beta(\text{opt}^{\ell_\infty}(A))^2.$$

Recall that the mechanism \mathcal{M}_g^{PA} is of the form $Ax + w$ where w is a noise vector whose distribution is independent of x . Thus the pair (\mathcal{D}, P) can be computed without looking at the database x . Therefore, the mechanism \mathcal{M}_{we} that samples a \mathcal{M} from \mathcal{D} and runs it on x is itself (ε, δ) -DP. The theorem follows. ■

Using the above result, we can now construct a mechanism that has small expected worse case error.

Theorem 24 *Let $A \in \mathbb{R}^{d \times N}$, ε, δ be given. There is an (ε, δ) -DP mechanism \mathcal{M}_{ew} , and a non-negative diagonal matrix P with $\text{tr}(P^2) = 1$ such that*

$$(\text{err}_{\mathcal{M}_{ew}}^{\ell_\infty}(A))^2 \leq O(\log^4 d \log((\log d)/\delta)) \cdot \text{specLB}(PA) \leq O(\log^4 d \log((\log d)/\delta)) \cdot (\text{opt}^{\ell_\infty}(A))^2$$

Proof: Our mechanism with small expected ℓ_∞ noise simply runs $L \triangleq O(\log d)$ copies $\mathcal{M}_{we}(x)$ and returns the median answer for each coordinate. In other words, if y^j denotes the outcome of the j th run of the mechanism \mathcal{M}_{we} , then we set $z_i = \text{median}(y_i^1, \dots, y_i^L)$.

By Markov's inequality, we know that for each i, j , $\Pr[(y_i^j - (Ax)_i)^2 \geq kE_i^{\mathcal{M}_{ew},A}] \leq \frac{1}{k}$. Applying Chernoff bounds, we conclude that $\Pr[(z_i - (Ax)_i)^2 \geq kE_i^{\mathcal{M}_{ew},A}] \leq (2/k)^{CL}$ for a constant C . By a union bound, we conclude that $\Pr[\|z - Ax\|_\infty \geq O(\log d) \text{opt}^{\ell_\infty}(A)] \leq d \cdot (\frac{2}{k})^{CL}$. Taking $L = \Omega(\log d)$ suffices to ensure that the integral of this probability over $k \in [4, \infty)$ converges, giving a bound on the expected ℓ_∞ norm of the error.

The resulting mechanism is not necessarily (ε, δ) -differentially private any more. However, since its outcome can be computed by postprocessing the outcome of L (ε, δ) -differentially private mechanisms, it is still $(L\varepsilon, L\delta)$ -differentially private. Scaling ε and δ by a factor of L , and substituting for β , we get the result. ■

6.2 Approximating Hereditary Discrepancy

In this section we show that our optimal dense case mechanism implies a polylogarithmic approximation to hereditary discrepancy. In particular, the mechanism optimal for ℓ_2^2 error can be used to approximate ℓ_2 discrepancy, and the mechanism optimal for ℓ_∞ error can be used to approximate ℓ_∞ discrepancy. The ℓ_∞ version of hereditary discrepancy is NP-hard to approximate to within a factor of $3/2$ (proof in appendix). Showing supeconstant hardness for approximating any version of hereditary discrepancy, constant hardness for approximating ℓ_2 hereditary discrepancy, and determining whether hereditary discrepancy can be exactly computed in nondeterministic polynomial time remain open problems.

Muthukrishnan and Nikolov [MN12] show that hereditary discrepancy gives a lower bound on the error of any mechanism.

Theorem 25 ([MN12], Corollary 1) *There exist constants ε, δ such that the following holds: Let A be a $d \times N$ matrix and \mathcal{M} be an (ε, δ) -differentially private mechanism. Then*

$$\text{herdisc}^{\ell_\infty}(A) \leq O(\log N) \cdot \text{err}_{\mathcal{M}}^{\ell_\infty}(A)$$

Fix constants ε and δ satisfying Theorem 25. Thus the theorem implies that $\text{herdisc}^{\ell_\infty}(A) \leq O(\log N) \cdot \text{opt}^{\ell_\infty}(A)$. It is easy to see that for any positive semidefinite diagonal matrix P with $\text{tr}(P^2) = 1$, $(\text{herdisc}^{\ell_\infty}(A))^2 \geq \text{herdisc}^{\ell_2^2}(PA)$: indeed for any $S \subseteq [N]$, there exists a coloring z supported on S such that $\|Az\|_\infty \leq \text{herdisc}^{\ell_\infty}(A)$. Since $\|PAz\|_2^2 \leq \|Az\|_\infty^2$, the claim follows. Moreover, recall that $\text{herdisc}^{\ell_2^2}(PA) \geq c \cdot \text{specLB}(PA)$, for some absolute constant c . Thus

$$\text{specLB}(PA) \leq (\text{herdisc}^{\ell_\infty}(A))^2$$

On the other hand, the mechanism of Theorem 24 satisfies

$$(\text{err}_{\mathcal{M}_{ew}}^{\ell_\infty}(A))^2 \leq O(\log^4 d \log((\log d)/\delta)) \text{specLB}(PA).$$

Thus we have

$$\begin{aligned} (\text{herdisc}^{\ell_\infty}(A))^2 &\leq O(\log^2 N) \text{err}_{\mathcal{M}_{ew}}^{\ell_\infty}(A) \\ &\leq O(\log^4 d \log^2 N \log((\log d)/\delta)) \cdot \text{specLB}(PA) \end{aligned}$$

We have sandwiched $\text{herdisc}^{\ell_\infty}(A)^2$ between two efficiently computable quantities that are $O(\log^4 d \log^2 N \log \log d)$ apart. It follows that

Theorem 26 *There is a polynomial time algorithm that, given a $d \times N$ real matrix A , outputs an $O(\log^2 d \log N \sqrt{\log \log d})$ approximation to $\text{herdisc}^{\ell_\infty}(A)$.*

The description of this mechanism \mathcal{M}_{ew} (which is specified by a distribution over $O(\log d)$ Gaussian noise addition mechanisms) thus serves as an efficiently computable and verifiable witness to an upper bound on the hereditary discrepancy of A .

We next give a constructive version of this result, by appealing to a result of Bansal [Ban10] that shows that the discrepancy of any set system can be constructively upper bounded by $O(\log dN)$ times the hereditary vector discrepancy. The vector discrepancy of a matrix A , denoted $\text{vecdisc}(A)$ is defined as the smallest λ such that the following semidefinite program is feasible:

$$\begin{aligned} \text{SDP VecDisc: } A &\in \mathbb{R}^{d \times m} \\ a^{(i)} V a^{(i)T} &\leq \lambda & \forall i \in [d] \\ V_{jj} &\leq 1 & \forall j \in [m] \\ \sum_{j=1}^m V_{jj} &\geq m/8 \\ V &\succeq 0 \end{aligned} \tag{20}$$

The hereditary vector discrepancy is simply $\text{hvecdisc}(A) \triangleq \max_{S \subseteq [N]} \text{vecdisc}(A|_S)$. We will show that given the mechanism \mathcal{M}_{ew} of Theorem 24, we can construct for any S a solution to the SDP corresponding to S . We note that the above SDP is a slight relaxation of the SDP used by Bansal: instead of the constraint $V_{jj} = 1$ for all j , we simply require each V_{jj} to be at most one, and that the trace of V is $\Omega(m)$. It is easy to verify that [Ban10] implies that:

Theorem 27 ([Ban10]) *Suppose that for a matrix $A \in \mathbb{R}^{d \times N}$ and a $\lambda \geq 0$, for every $S \subseteq [N]$ with $\text{rank}(A|_S) = |S|$, it is the case that the SDP 20 is feasible for $A|_S$. Then there is polynomial time algorithm that finds a coloring χ of $[N]$ with discrepancy at most $O(\lambda \log dN)$.*

We consider a variant of the above SDP, where we drop the $V_{jj} \leq 1$ constraint and require the trace of V to be slightly larger:

$$\begin{aligned} \text{SDP RelVecDisc: } A &\in \mathbb{R}^{d \times m} \\ a^{(i)} V a^{(i)T} &\leq \lambda & \forall i \in [d] \\ \sum_{j=1}^m V_{jj} &\geq m \\ V &\succeq 0 \end{aligned} \tag{21}$$

We first show that it suffices to satisfy the SDP 21.

Lemma 16 *Suppose that for a matrix $A \in \mathbb{R}^{d \times N}$, for every $S \subseteq [N]$ with $\text{rank}(A|_S) = |S|$, it is the case that the SDP 21 is feasible for $A|_S$ and λ . Then for any $S \subseteq [N]$ with $\text{rank}(A|_S) = |S|$, the SDP 20 is feasible for $A|_S$ and 2λ .*

Proof: We construct a feasible solution to 20 by repeatedly using solutions to 21 for different values of the restriction $A|_S$. Fix an $S \subseteq [N]$ with $|S| = m$ and let $S_0 = S$. Let W^0 be the $d \times m$ zero matrix indexed by $S \subseteq [N]$. Let V^0 be a solution to SDP 21 for $A|_{S_0}$. Let $\gamma_0 = \max_j V_{jj}^0$, and let $\bar{V}^0 = V^0/2\gamma_0$. For each $j, j' \in S_0$, set $W_{jj'}^1 = W_{jj'}^0 + \bar{V}_{jj'}^0$. Finally we set $S_1 = S_0 \setminus \{j : W_{jj}^1 \geq \frac{1}{4}\}$.

Given S_i such that $|S_i| \geq m/2$, we let V^i be a feasible solution to the SDP for $A|_{S_i}$ and set $\gamma_i = \max_j V_{jj}^i$, and $\bar{V}^i = V^i/2\gamma_i$. For each $j, j' \in S_i$, set $W_{jj'}^{i+1} = W_{jj'}^i + \bar{V}_{jj'}^i$. We update $S_{i+1} = S_i \setminus \{j : W_{jj}^{i+1} \geq \frac{1}{4}\}$. We stop once $|S_i|$ falls below $m/2$, say in iteration L . It is easy to see that we delete at least one j from S_i in each step, so that this process converges.

We claim that W^L is a feasible solution to SDP 20. Observe that W^L is simply a non-negative linear combination of V^i 's and hence is positive semidefinite. By definition of \bar{V} , each diagonal entry is at most $\frac{1}{2}$, so that the definition of S_i ensures that $W_{jj}^L \leq \frac{3}{4} < 1$ for all j . Moreover, for each $j \in S \setminus S_L$, the entry $W_{jj}^L \geq \frac{1}{4}$, and there are at least $m/2$ such j , so that the trace of W^L is at least $m/8$ as required. Finally, $a^{(i)}W^La^{(i)T}$ is simply the sum $\sum_{t=0}^{L-1} \frac{a^{(i)}V^ta^{(i)T}}{2\gamma_t}$. Thus it suffices to show that $\sum_{t=1}^{L-1} (2\gamma_t)^{-1} \leq 2$. Note that $\text{tr}(W^{t+1} - W^t) = \text{tr}(\bar{V}^t) \geq (2\gamma_t)^{-1}|S_t| \geq (2\gamma_t)^{-1}m/2$. It follows that $m \geq \text{tr}(W^L) \geq \sum_{t=1}^{L-1} (2\gamma_t)^{-1}m/2$ so that $\sum_{t=1}^{L-1} (2\gamma_t)^{-1} \leq 2$ as needed. The claim follows. ■

Finally, we describe how to use the mechanism \mathcal{M}_{ew} of Theorem 24 to construct a feasible solution to SDP 21.

Lemma 17 *Suppose that for a matrix $A \in \mathbb{R}^{d \times N}$, we have an oblivious noise (ε, δ) -DP mechanism \mathcal{M} such that $\text{err}_{\mathcal{M}}^{\ell_\infty}(A) \leq \kappa$. Then for any $S \subseteq [N]$ with $\text{rank}(A|_S) = |S|$, the SDP 21 is feasible for $A|_S$ with $\lambda = O_{\varepsilon, \delta}(\kappa^2)$.*

Proof: Since an (ε, δ) -DP mechanism for A also yields an (ε, δ) -DP mechanism for any submatrix of A , we can assume without loss of generality that $S = [N]$. For a $y \in \mathbb{R}^d$, let $\text{inv } A(y) \triangleq \arg \min_x \|Ax - y\|_\infty$. Let Y be a random variable distributed according to $\mathcal{M}(0)$ and let $X = \text{inv } A(Y)$. Then by the properties of the mechanism $\mathbb{E}[\|AX\|_\infty] \leq \mathbb{E}[\|AX - Y\|_\infty + \|Y\|_\infty] \leq 2\mathbb{E}[\|Y\|_\infty] \leq 2\kappa$, since 0 is a possible value for $\text{inv } A(Y)$, and we picked X . Thus if we set $V = E[XX^T]$, then $a^{(i)}Va^{(i)T} \leq 16\kappa^2$ for each i .

It remains to lower bound $V_{jj} = \mathbb{E}[(X_j)^2] \geq \text{Var}(X_j)$. Intuitively, if $\text{Var}(X_j)$ was small, then Y gives too much information about X_j , violating differential privacy. Formally, by differential privacy, the statistical distance between $\mathcal{M}(0)$ and $\mathcal{M}(e_j)$ is at most $2(\varepsilon + \delta)$. Thus the distribution Y is close to $Y + Ae_j$. But it is easy to check² that $\text{inv } A(y + Ae_j) = \text{inv } A(y) + e_j$. Thus the distribution X is close to $X + e_j$. This in turn lower bounds the variance of X_j by an absolute constant. Scaling V by a constant factor implies the result. Finally, we observe that this proof can be easily made constructive. We can take polynomially many samples x_1, \dots, x_k and setting V to the empirical estimate $\frac{1}{k} \sum_{i=1}^k x_i x_i^T$ instead of $E[XX^T]$. Truncating the distribution so as to reject any x_i with $\|Ax_i\|_\infty \geq d^3\kappa$ or with $\|x_i\|_\infty \geq d^3$ still preserves the required properties and allows us to use Chernoff bounds to ensure that the empirical estimate satisfies the constraints up to a constant factor. ■

7 Conclusions

We have presented near optimal mechanisms for any linear query for dense and sparse databases, under both pure and approximate differential privacy. Our mechanisms are simple and efficient, and it would be instructive to implement them so as to compare them with existing techniques.

Our work uses the hereditary discrepancy lower bound, which holds for small enough constant ε and δ . Since our lower bounds do not get higher as δ gets smaller, the approximation ratio has an $O(\sqrt{\log 1/\delta})$ term in it. We leave open the question of developing better lower bounding techniques, and better approximation ratios for (ε, δ) -differentially private mechanisms. Our work gives ℓ_2^2 bounds on the error. While we can translate those bounds to ℓ_∞ error bounds for the dense case, we leave open the question of designing near optimal ℓ_∞ error mechanisms in the sparse case.

8 Acknowledgements

The first and second named authors would like to thank Moritz Hardt and Daniel Dadush for several helpful discussions. The first named author was partially supported by a Simons Graduate Fellowship, award number 252861.

²We assume that ties in the definition of $\text{inv } A(\cdot)$ are broken at random.

References

- [AHK12] Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta algorithm and applications. *Theory of Computing*, 8:121–164, 2012.
- [AS93] I. Adler and R. Shamir. A randomized scheme for speeding up algorithms for linear and convex programming problems with high constraints-to-variables ratio. *Mathematical programming*, 61(1):39–52, 1993.
- [Ban10] N. Bansal. Constructive algorithms for discrepancy minimization. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 3–10. IEEE, 2010.
- [BCD⁺07] Boaz Barak, Kamalika Chaudhuri, Cynthia Dwork, Satyen Kale, Frank McSherry, and Kunal Talwar. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In Leonid Libkin, editor, *Proceedings of ACM PODS*, pages 273–282. ACM, 2007.
- [BDKT12] Aditya Bhaskara, Daniel Dadush, Ravishankar Krishnaswamy, and Kunal Talwar. Unconditional differentially private mechanisms for linear queries. In *Proceedings of the 44th symposium on Theory of Computing*, STOC ’12, pages 1269–1284, New York, NY, USA, 2012. ACM.
- [BDMN05] A. Blum, C. Dwork, F. McSherry, and K. Nissim. Practical privacy: the sulq framework. In *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 128–138. ACM, 2005.
- [BF88] I. Bárány and Z. Füredi. Approximation of the sphere by polytopes having few vertices. *Proceedings of the American Mathematical Society*, 102(3):651–659, 1988.
- [BLR08] Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to non-interactive database privacy. In *STOC ’08: Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 609–618, New York, NY, USA, 2008. ACM.
- [BN10] Hai Brenner and Kobbi Nissim. Impossibility of differentially private universally optimal mechanisms. In *Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, FOCS ’10, pages 71–80, Washington, DC, USA, 2010. IEEE Computer Society.
- [Bor75] C. Borell. The brunn-minkowski inequality in gauss space. *Inventiones Mathematicae*, 30(2):207–216, 1975.
- [BS95] József Beck and Vera T. Sós. Handbook of combinatorics (vol. 2). chapter Discrepancy theory, pages 1405–1446. MIT Press, Cambridge, MA, USA, 1995.
- [BT87] J. Bourgain and L. Tzafriri. Invertibility of large submatrices with applications to the geometry of banach spaces and harmonic analysis. *Israel journal of mathematics*, 57(2):137–224, 1987.
- [Cha00] Bernard Chazelle. *The Discrepancy Method: Randomness and Complexity*. Cambridge University Press, 2000.
- [Cla10] K.L. Clarkson. Coresets, sparse greedy approximation, and the frank-wolfe algorithm. *ACM Transactions on Algorithms (TALG)*, 6(4):63, 2010.
- [CNN11] M. Charikar, A. Newman, and A. Nikolov. Tight hardness results for minimizing discrepancy. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1607–1614. SIAM, 2011.
- [CSS10] T-H. Hubert Chan, Elaine Shi, and Dawn Song. Private and continual release of statistics. In *ICALP*, 2010.
- [CV11] Karthekeyan Chandrasekaran and Santosh Vempala. A discrepancy based approach to integer programming. *CoRR*, abs/1111.4649, 2011.
- [De12] A. De. Lower bounds in differential privacy. *Theory of Cryptography*, pages 321–338, 2012.

- [DKM⁺06a] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Proc. 25th EUROCRYPT*, pages 486–503. Springer, 2006.
- [DKM⁺06b] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation, 2006.
- [DMNS06] C. Dwork, F. Mcsherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, 2006.
- [DMT07] Cynthia Dwork, Frank McSherry, and Kunal Talwar. The price of privacy and the limits of LP decoding. In *Proc. 39th STOC*, pages 85–94. ACM, 2007.
- [DN03] Irit Dinur and Kobbi Nissim. Revealing information while preserving privacy. In *Proc. 22nd PODS*, pages 202–210. ACM, 2003.
- [DNR⁺09] Cynthia Dwork, Moni Naor, Omer Reingold, Guy N. Rothblum, and Salil Vadhan. On the complexity of differentially private data release: efficient algorithms and hardness results. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, STOC '09, pages 381–390, New York, NY, USA, 2009. ACM.
- [DRV10] Cynthia Dwork, Guy N. Rothblum, and Salil Vadhan. Boosting and differential privacy. In *Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, FOCS '10, pages 51–60, Washington, DC, USA, 2010. IEEE Computer Society.
- [DWHL11] Bolin Ding, Marianne Winslett, Jiawei Han, and Zhenhui Li. Differentially private data cubes: optimizing noise sources and consistency. In *SIGMOD Conference*, pages 217–228, 2011.
- [DY08] Cynthia Dwork and Sergey Yekhanin. New efficient attacks on statistical disclosure control mechanisms. In *Proc. 28th CRYPTO*, pages 469–480. Springer, 2008.
- [FMN] Nadia Fawaz, S. Muthukrishnan, and Aleksandar Nikolov. Nearly optimal private convolutions. unpublished manuscript.
- [FW56] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2):95–110, 1956.
- [GHRU11] Anupam Gupta, Moritz Hardt, Aaron Roth, and Jonathan Ullman. Privately releasing conjunctions and the statistical query barrier. In *STOC*, pages 803–812, 2011.
- [Gia03] Apostolos Giannopoulos. Notes on isotropic convex bodies. Available at <http://users.uoa.gr/~apgiannop/notes.html>, 2003.
- [Glu07] E.D. Gluskin. Extremal properties of orthogonal parallelepipeds and their applications to the geometry of banach spaces. *Mathematics of the USSR-Sbornik*, 64(1):85, 2007.
- [GRS09] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. In *STOC*, pages 351–360, 2009.
- [GRU12] Anupam Gupta, Aaron Roth, and Jonathan Ullman. Iterative constructions and private data release. In *TCC*, pages 339–356, 2012.
- [GS10] Mangesh Gupte and Mukund Sundararajan. Universally optimal privacy mechanisms for minimax agents. In *Proceedings of the twenty-ninth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, PODS '10, pages 135–146, New York, NY, USA, 2010. ACM.
- [Gur00] V. Guruswami. Inapproximability results for set splitting and satisfiability problems with no mixed clauses. *Approximation Algorithms for Combinatorial Optimization*, pages 155–166, 2000.
- [HLM12] Moritz Hardt, Katrina Ligett, and Frank McSherry. A simple and practical algorithm for differentially private data release. In *NIPS*, 2012. To appear.
- [HR10] Moritz Hardt and Guy N. Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, FOCS '10, pages 61–70, Washington, DC, USA, 2010. IEEE Computer Society.

- [HRMS10] Michael Hay, Vibhor Rastogi, Gerome Miklau, and Dan Suciu. Boosting the accuracy of differentially private histograms through consistency. *PVLDB*, 3(1):1021–1032, 2010.
- [HT10] Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Proceedings of the 42nd ACM symposium on Theory of computing*, STOC '10, pages 705–714, New York, NY, USA, 2010. ACM.
- [Joh48] F. John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays presented to R. Courant on his 60th Birthday*, pages 187–204, 1948.
- [Kha96] L.G. Khachiyan. Rounding of polytopes in the real number model of computation. *Mathematics of Operations Research*, 21(2):307–320, 1996.
- [KRS13] Shiva Kasiviswanathan, Mark Rudelson, and Adam Smith. The power of linear reconstruction attacks. In *SODA*, 2013. To appear.
- [KRSU10] Shiva Prasad Kasiviswanathan, Mark Rudelson, Adam Smith, and Jonathan Ullman. The price of privately releasing contingency tables and the spectra of random matrices with correlated rows. In *Proceedings of the 42nd ACM symposium on Theory of computing*, STOC '10, pages 775–784, New York, NY, USA, 2010. ACM.
- [KY05] P. Kumar and EA Yildirim. Minimum-volume enclosing ellipsoids and core sets. *Journal of Optimization Theory and Applications*, 126(1):1–21, 2005.
- [Lar11] Kasper Green Larsen. On range searching in the group model and combinatorial discrepancy. In *FOCS*, pages 542–549, 2011.
- [LHR⁺10] Chao Li, Michael Hay, Vibhor Rastogi, Gerome Miklau, and Andrew McGregor. Optimizing linear counting queries under differential privacy. In *Proceedings of the twenty-ninth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, PODS '10, pages 123–134, New York, NY, USA, 2010. ACM.
- [LM12a] Chao Li and Gerome Miklau. An adaptive mechanism for accurate query answering under differential privacy. *PVLDB*, 5(6):514–525, 2012.
- [LM12b] Chao Li and Gerome Miklau. Measuring the achievable error of query sets under differential privacy. *CoRR*, abs/1202.3399, 2012.
- [LSV86] L. Lovász, J. Spencer, and K. Vesztergombi. Discrepancy of set-systems and matrices. *European Journal of Combinatorics*, 7(2):151–160, 1986.
- [Mat99] Jiří Matoušek. *Geometric Discrepancy (An Illustrated Guide)*. Springer, 1999.
- [Mat11] Jiří Matoušek. The determinant bound for discrepancy is almost tight. <http://arxiv.org/abs/1101.0767>, 2011.
- [MN12] S. Muthukrishnan and Aleksandar Nikolov. Optimal private halfspace counting via discrepancy. In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, pages 1285–1292, New York, NY, USA, 2012. ACM.
- [NRS07] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84, New York, NY, USA, 2007. ACM.
- [RHS07] Vibhor Rastogi, Sungho Hong, and Dan Suciu. The boundary between privacy and utility in data publishing. In *VLDB*, pages 531–542, 2007.
- [RR10] Aaron Roth and Tim Roughgarden. Interactive privacy via the median mechanism. In *Proceedings of the 42nd ACM symposium on Theory of computing*, STOC '10, pages 765–774, New York, NY, USA, 2010. ACM.
- [RWY11] G. Raskutti, M.J. Wainwright, and B. Yu. Minimax rates of estimation for high-dimensional linear regression over ℓ_1 formula formulatype=. *Information Theory, IEEE Transactions on*, 57(10):6976–6994, 2011.

- [Sch78] T.J. Schaefer. The complexity of satisfiability problems. In *Proceedings of the tenth annual ACM symposium on Theory of computing*, pages 216–226, 1978.
- [Tao12] Terence Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.
- [TY07] M.J. Todd and E.A. Yildirim. On khachiyan’s algorithm for the computation of minimum-volume enclosing ellipsoids. *Discrete Applied Mathematics*, 155(13):1731–1744, 2007.
- [Ver01] R. Vershynin. John’s decompositions: Selecting a large part. *Israel Journal of Mathematics*, 122(1):253–277, 2001.
- [XWG10] Xiaokui Xiao, Guozhang Wang, and Johannes Gehrke. Differential privacy via wavelet transforms. In *ICDE*, pages 225–236, 2010.
- [XXY10] Yonghui Xiao, Li Xiong, and Chun Yuan. Differentially private data release through multidimensional partitioning. In *Secure Data Management*, pages 150–168, 2010.
- [YZW⁺12] Ganzhao Yuan, Zhenjie Zhang, Marianne Winslett, Xiaokui Xiao, Yin Yang, and Zhifeng Hao. Low-rank mechanism: Optimizing batch queries under differential privacy. *PVLDB*, 5(11):1352–1363, 2012.

A Concentration of Log concave measures

The following measure concentration inequality is standard. We include a proof below for completeness. We start by stating the Brunn-Minkowski inequality.

Theorem 28 *Let μ be a log-concave measure on \mathbb{R}^d , let $\alpha, \beta \geq 0$ be numbers such that $\alpha + \beta = 1$, and let $A, B \subseteq \mathbb{R}^d$ be measurable sets such that the set $\alpha A + \beta B$ is measurable. Then*

$$\mu(\alpha A + \beta B) \geq (\mu(A))^\alpha (\mu(B))^\beta$$

This can be used to prove Borell’s lemma for arbitrary log concave distributions. We use the proof approach presented in [Gia03].

Proof of Theorem 20: Observe that

$$\mathbb{R}^d \setminus A \supseteq \left(\frac{2}{t+1}\right)(\mathbb{R}^d \setminus tA) + \frac{t-1}{t+1}A.$$

Applying the Brunn-Minkowski inequality and rearranging proves the result. ■

B From Concentration to Expectation

We repeatedly use the fact that a exponential or better bound on the upper tail implies a bound on the expectation. For completeness, we give a quantitative version with a proof.

Lemma 18 *Suppose that for some constants C, c_1, c_2 , a random variable X satisfies: $\forall \alpha \geq c_1, \Pr[X \geq \alpha C] \leq \exp(-\alpha/c_2)$. Then $\mathbb{E}[X] \leq (c_1 + c_2)C$.*

Proof: Let $Y = (X - c_1 C)/C$. Then

$$\begin{aligned} E[Y] &= \int_0^\infty \Pr[Y \geq \alpha] d\alpha \\ &\leq \int_0^\infty \exp(-\alpha/c_2) d\alpha \\ &= c_2 \end{aligned}$$

The claim follows by linearity of expectation. ■

C Hardness of Approximating Hereditary Discrepancy

We show constant hardness of approximation $\text{herdisc}^{\ell_\infty}$. Strong inapproximability results were previously given for discrepancy, in the ℓ_2 and ℓ_∞ versions, by Charikar, Newman, and Nikolov [CNN11].

Let us define

$$\text{disc}^{\ell_\infty}(A) = \min_{x \in \{0,1\}^n} \|Ax\|_\infty$$

Theorem 29 *There exists a family of input matrices $A \in \{0,1\}^{m \times n}$ for which it is NP-hard to distinguish between the two cases (1) $\text{herdisc}^{\ell_\infty}(A) \leq 2$ and (2) $\text{disc}^{\ell_\infty}(A) \geq 3$.*

The proof of Theorem 29 is a straight-forward reduction from the 2-colorability problem for 3-uniform hypergraphs. A maximization version of this problem (i.e. maximize the number of bichromatic edges) is also known as **Max-E3-Set Splitting** and is equivalent to **NotAllEqual-SAT** restricted to inputs with no negated variables.

Definition 2 *A hypergraph $H = (V, E)$, where $E \subseteq 2^V$, is 2-colorable if and only if there exists a set $T \subseteq V$ such that for all $e \in E$, $T \cap e \neq e$ and $T \cap e \neq \emptyset$. The set T is called a transversal of H .*

Lemma 19 ([Sch78]) *There exists a family of 3-uniform hypergraphs such that deciding whether a hypergraph in the family is 2-colorable is NP-complete.*

Proof of Theorem 29: The reduction simply maps a 3-uniform hypergraph to its incidence matrix. I.e. for a hypergraph $H = (V, E)$, where $V = \{v_1, \dots, v_n\}$ and $E = \{e_1, \dots, e_m\}$, we create a $m \times n$ matrix A , where $A_{ij} = 1$ if $v_j \in e_i$ and $A_{ij} = 0$ otherwise. Observe that if H is 2-colorable, and this is witnessed by a transversal T , then $\|(A|_S)x\|_\infty \leq 2$ for all $S \subseteq [m]$ and for x defined by $x_i = +1 \Leftrightarrow v_i \in T$. On the other hand, if H is not 2-colorable, for any $x \in \{+1, -1\}^n$ we have $\|Ax\|_\infty \geq 3$, since otherwise the set $T = \{v_i : x_i = +1\}$ would be a transversal. ■

We note that Guruswami proved constant hardness of approximation for **Max-E3-Set Splitting** [Gur00]. In particular, he showed that it is NP-hard to distinguish 2-colorable 3-uniform hypergraphs from hypergraphs for which any coloring with 2 colors leaves at least a $1/20$ fraction of the edges monochromatic.